


```

VERSION      AI275105.1  GI:3897379
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 451)
              NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
              National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
              Unpublished (1997)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs@mail.nih.gov
              This clone is available royalty-free through LNL ; contact the
              IMAGE Consortium (info@image.lnl.gov) for further information.
              Insert Length: 932 Std Error: 0.00
              Seq primer: -40UP from Gibco.
FEATURES     Location/Qualifiers
              1..451
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:1877205"
               /tissue type="Pooled human melanocyte, fetal heart, and
               pregnant uterus"
               /lab host="DH10B"
               /clone lib="Scars NHMPu SI"
               /note="Organ: mixed (see below); Vector: pTT3D-Pac
               (Pharmacia) with a modified polylinker; Site 1: Not 1;
               Site 2: Eco RI; Equal amounts of plasmid DNA from three
               normalized libraries (melanocyte 2NBHM, pregnant uterus
               NBHPU, and fetal heart NBH19W) were mixed, and ss circles
               were made in vitro. Following HAP purification, this DNA
               was used as tracer in a subtractive hybridization
               reaction. The driver was PCR-amplified cDNAs from pools of
               5,000 clones made from the same 3 libraries. The pools
               consisted of I.M.A.G.E. clones 260232-265223,
               340488-345479, and 484488-489479."
ORIGIN
Query Match      18.1%; Score 450; DB 9; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.5e-76;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2035 TTCCCCATCCCTTCTTCTCTCATGGTACGTGCTCTCTTAATATAGCGTTGGTGAGA 2094
DB 451 TTCCCCATCCCTTCTTCTCTCATGGTACGTGCTCTCTTAATATAGCGTTGGTGAGA 392

QY 2095 TTTTCAGTGGTCCATATTCCTTCCCTCTGTTGCTTTCCTGAGTATCCACTAAG 2154
DB 391 TTTTCAGTGGTCCATATTCCTTCCCTCTGTTGCTTTCCTGAGTATCCACTAAG 332

QY 2155 AATATTTTGTGTTCTTTCTCAGGGAATCTAAGGGAGGAATATCAACTGTGCACAAG 2214
DB 331 AATATTTTGTGTTCTTTCTCAGGGAATCTAAGGGAGGAATATCAACTGTGCACAAG 272

QY 2215 GAAAAAATAGATATGTGAAGTTTCAGTAAATTTCTCACAATCAGAGATTAAT 2274
DB 271 GAAAAAATAGATATGTGAAGTTTCAGTAAATTTCTCACAATCAGAGATTAAT 212

QY 2275 TCAGAAAGGAGAAAAACACAGACCAAGAGAGTATCTAAGACCAGGAGTGTGTTTAT 2334
DB 211 TCAGAAAGGAGAAAAACACAGACCAAGAGAGTATCTAAGACCAGGAGTGTGTTTAT 152

QY 2335 TAATGCTAGGATGAAGAAATGCATAGAACATTTAGTACTTTGTAATTAATCAGAAATAA 2394
DB 151 TAATGCTAGGATGAAGAAATGCATAGAACATTTAGTACTTTGTAATTAATCAGAAATAA 92

QY 2395 CATGATTTAGTCATATTTGTGAAAAATTAATATAATTTTCTTGATTTTGTCTGTAT 2454
DB 91 CATGATTTAGTCATATTTGTGAAAAATTAATATAATTTTCTTGATTTTGTCTGTAT 32

QY 2455 CTGTGAAAAAATAAATTTCTTTATAAAACTC 2484

```

```

DB 31 CTGTGAAAAAATAAATTTCTTTATAAAACTC 2

RESULT 10
LOCUS   AL541045/c
DEFINITION AL541045 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YC19
3-PRIME, mRNA sequence.
ACCESSION AL541045
VERSION   AL541045.2  GI:30544837
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS   Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE     Pull-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 15, 2001 this sequence version replaced gi:12871741.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1009.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE005AB10NP1&cluster=1009.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DE005AB10NP1.
FEATURES     Location/Qualifiers
              1..1201
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="CS0DE005YC19"
               /tissue type="PLACENTA"
               /clone lib="Homo sapiens PLACENTA"
               /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
               with a NotI-oligo(dT) primer. Five prime end enriched,
               double-strand cDNA was digested with Not I and cloned into
               the Not I and EcoRV sites of the pCMVSPORT 6 vector.
               Library was not normalized."
ORIGIN
Query Match      17.8%; Score 442.6; DB 9; Length 1201;
Best Local Similarity 90.5%; Pred. No. 3.4e-75;
Matches 448; Conservative 19; Mismatches 28; Indels 0; Gaps 0;

QY 801 AAATATTTTGAATTAATATCTTGCTTTATGTCTACCTTACAGTTTATCCCATGGAAT 860
DB 791 ACATATTTCTATATCTTTGAACGGTAACATTCGTGATAGATACCAGTTTATCCCATGGAAT 732

QY 861 CAATCAAAATAATCATCATGACTAGTATGATTCCTCCATCGCTGAAGGACAGATTCAT 920
DB 731 CAATCAAAATAATCATCATGACTAGTATGATTCCTCCATCGCTGAAGGACAGATTCAT 672

QY 921 GTGTGGCATTTGATTTGATGCGAGCTCTATTCAATACCTTCTCTCAGATGATAGTAA 980
DB 671 GTGTGGCATTTGATTTGATGCGAGCTCTATTCAATACCTTCTCTCAGATGATAGTAA 612

QY 981 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGCTGGTGTGTAATGTTGGCTTCTCA 1040
DB 611 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGCTGGTGTGTAATGTTGGCTTCTCA 552

QY 1041 CTCATGTGATAGCATTGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGG 1100
DB 551 CTCATGTGATAGCATTGGATTTGATTACAAAAGGTGACCTTATAGAAATAGAGAGATGG 492

QY 1101 AGCCTGTGAGGTCCCAAGCTAGAGGAAGTCCAAAGAAAATCTTGATTTGCTTTCTGACA 1160

```

```

Db 491 CGCTGTGCGTTCAGCTGAGGAGTCCAGAAATCTGATTTGCTCTTCGACM 432
Qy 1161 TCTCGGTTAGCAATATCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 1220
Db 431 TCTCGGTTAGCAATATCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 372
Qy 1221 TTCTTTCTGCTGAGACGAATCTATGGGCTCAGATGACTTCTAGAGGATTCGCTT 1280
Db 371 TTCTTTCTGCTGAGACGAATCTATGGGCTCAGATGACTTCTAGAGGATTCGCTT 312
Qy 1281 TTGAGCAATAGGTA 1295
Db 311 TTGAGCAATAGGTA 297

RESULT 11
AW130518/c
LOCUS xF52408.x1 NCI_CGAP_Gas4 452 bp mRNA linear EST 27-OCT-1999
DEFINITION mRNA sequence.
ACCESSION AW130518
VERSION AW130518.1 GI:6132125
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-x@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 420.
FEATURES
source
1..452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2621679"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH103"
/clone_lib="NCI_CGAP Gas4"
/notes="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

Query Match 17.7%; Score 440.8; DB 10; Length 452;
Best Local Similarity 99.5%; Pred. No. 9e-75;
Matches 442; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2041 CATCCCTCTCTCCCTCATGACGTGCTCTCTAATATAGCGTTGGTTGAGATTTTCA 2100
Db 452 CATCCCTCTCTCCCTCATGACGTGCTCTCTAATATAGCGTTGGTTGAGATTTTCA 393
Qy 2101 GTGGTCCAATATCTCTCTCCCTCTCGGTTGCCCTTCCCTGAGATAATCCACTAAGAAATATT 2160
Db 392 GTGGTCCAATATCTCTCTCCCTCTCGGTTGCCCTTCCCTGAGATAATCCACTAAGAAATATT 333
```

```

Qy 2161 TTGTGTTTCTTTCTCAGGGAAATCTAAGGAGGAAATATATCACTGTGCAAGGAAAAA 2220
Db 332 TTGTGTTTCTTTCTCAGGGAAATCTAAGGAGGAAATATATCACTGTGCAAGGAAAAA 273
Qy 2221 AATAGATATGTGAAGGTTCAAGTAAATTTCCCTCACATCACAGAAAGATTAAATTCAGAA 2280
Db 272 AATAGATATGTGAAGGTTCAAGTAAATTTCCCTCACATCACAGAAAGATTAAATTCAGAA 213
Qy 2281 AGGAGAAAAACACAGACCAAGAGAGATATCTAGACCAAGAGGATGTGTTTATTAATGT 2340
Db 212 AGAAGAAAAACACAGACCAAGAGAGATATCTAGACCAAGAGGATGTGTTTATTAATGT 153
Qy 2341 CTAGGATGAAGAAATGCATAGAACATTTGTAGTACTTCTTAATAACTAGAAATAACATGAT 2400
Db 152 CTAGGATGAAGAAATGCATAGAACATTTGTAGTACTTCTTAATAACTAGAAATAACATGAT 93
Qy 2401 TTAGTCAATATGTGAAAAATATATAATTTTCTTGGATTTATGTCTGTCATCTGTGA 2460
Db 92 TTAGTCAATATGTGAAAAATATATAATTTTCTTGGATTTATGTCTGTCATCTGTGA 33
Qy 2461 AAAAAATAAATTTCTTATAAAACTC 2484
Db 32 AAAAAATAAATTTCTTATAAAACTC 9

RESULT 12
AW1290106/c
LOCUS QW26G02.x1 NCI_CGAP_Ut-4 436 bp mRNA linear EST 21-DEC-1998
DEFINITION mRNA sequence.
ACCESSION AW1290106
VERSION AW1290106.1 GI:3931772
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-x@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 765 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 405.
FEATURES
source
1..436
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1992195"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Ut-4"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"

Query Match 17.6%; Score 436; DB 9; Length 436;
Best Local Similarity 100.0%; Pred. No. 7.6e-74;
```


Db	121	TCCCTTCTCTTCTCATGTGCTCTCTAAATATTAGCGTTGGTTGAGATTTTCAGT	180
Qy	2103	GGTCCAAATATTCCTTTCCCTCTGGTTGCGCTTTCCTGAGATAATCCACTAAGAATAATTTT	2162
Db	181	GGTCCAAATATACCTTCTTCCCTCTGGTTGCGCTTTCCTGAGATAATCCACTAAGAATAATTTT	240
Qy	2163	GTGTTTCTTTTCTCAGGGAATCTAAGGAGGAGAAATATCAACTGTGCACAAGGAAAAAAA	2222
Db	241	GTGTTTCTTTTCTCAGGGAATCTAAGGAGGAGAAATATCAACTGTGCACAAGGAAAAAAA	300
Qy	2223	TAGATATGTGAAAGGTTTCAAGTAATTTCTTCACATCACAGAGATTAAAAATTCAGAAAG	2282
Db	301	TAGATATGTGAAAGGTTTCAAGTAATTTCTTCACATCACAGAGATTAAAAATTCAGAAAG	360
Qy	2283	GAGAAAAACAGACAGCCAA-AGAGAAGTATCTAAGACCAAGGAGTGTCTTTTATTAAT-GT	2340
Db	361	GAGAAAAANACAGACCAAGAGAGAAGTATCTAAGACCAAGGAGTGTCTTTTATTAATGGT	420
Qy	2341	CTAGAGTCAGAAATGCATA-GRACATTGTA-GTACTTGTATAAATCTAGAAATAACATG	2398
Db	421	CTAGATGAGANNATGCATTAGGAACATTGTAGGACTTTGTATAANTACTAGAAATACCTGG	480
Qy	2399	ATTTA-GTCATAATGTGAAAAA	2420
Db	481	NTTTAGGTCTATAATTTGTGAAAA	503

Search completed: April 26, 2004, 01:59:38
Job time : 4512.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 123.228 Seconds
(without alignments)
11186.557 Million cell updates/sec

Title: US-10-051-835-18
Perfect score: 2484
Sequence: 1 gtttatctgtattataac.....ataaattttataaaactc 2484

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	240	9.7	240	4	US-09-397-787-62
2	196.8	7.9	2494	4	US-09-566-921-125
3	90.6	3.6	7218	1	US-08-232-463-14
4	53.6	2.2	551	3	US-09-385-982-122
5	48	1.9	6306	4	US-10-204-708-49
6	47.8	1.9	5332	4	US-09-801-861-3
7	45.8	1.8	1353	4	US-09-601-198-37
8	45.2	1.8	660	1	US-07-991-867B-32
9	45.2	1.8	660	1	US-08-107-755A-32
10	45.2	1.8	660	2	US-08-544-332-32
11	45.2	1.8	660	4	US-09-370-861A-32
12	45.2	1.8	1511	1	US-07-991-867B-8
13	45.2	1.8	1511	1	US-08-107-755A-8
14	45.2	1.8	1511	2	US-08-544-332-8
15	45.2	1.8	1511	4	US-09-370-861A-8
16	45.2	1.8	4810	3	US-08-852-629-15
17	45.2	1.8	4838	3	US-08-852-629-11
18	45.2	1.8	7218	1	US-08-232-463-14
19	44.2	1.8	8961	4	US-10-204-708-80
20	44.2	1.8	20674	4	US-09-641-638-651
21	44.2	1.8	5219	4	US-10-204-708-52
22	43.8	1.8	19513	4	US-10-204-708-39
23	43.4	1.7	20674	4	US-09-641-638-651
24	43	1.7	832	4	US-09-621-976-2813
25	43	1.7	11049	4	US-10-204-708-22
26	42.8	1.7	832	4	US-09-621-976-2813
27	42.8	1.7	5217	4	US-09-100-703A-25

C 28	42.6	1.7	6020	4	US-10-204-708-8	Sequence 8, Appli
C 29	42.6	1.7	640681	4	US-09-790-988-1	Sequence 1, Appli
C 30	42.4	1.7	658	3	US-08-998-416-595	Sequence 595, App
C 31	42.4	1.7	6156	4	US-10-204-708-60	Sequence 60, Appl
C 32	42	1.7	8920	2	US-08-446-855A-1	Sequence 1, Appli
C 33	42	1.7	8920	3	US-09-150-741-1	Sequence 1, Appli
C 34	41.8	1.7	1667	1	US-08-485-284A-1	Sequence 1, Appli
C 35	41.8	1.7	148567	4	US-09-801-876B-3	Sequence 3, Appli
C 36	41.8	1.7	148567	4	US-10-254-869-3	Sequence 3, Appli
C 37	41.6	1.7	396	4	US-09-640-173-53	Sequence 53, Appl
C 38	41.6	1.7	396	4	US-09-713-550-53	Sequence 53, Appl
C 39	41.6	1.7	6182	4	US-10-204-708-87	Sequence 87, Appl
C 40	41.2	1.7	5915	4	US-10-204-708-77	Sequence 77, Appl
C 41	41.2	1.7	10467	4	US-10-204-708-2	Sequence 2, Appli
C 42	41	1.7	11131	4	US-10-204-708-28	Sequence 28, Appl
C 43	40.8	1.6	782	3	US-08-998-416-224	Sequence 224, App
C 44	40.8	1.6	640681	4	US-09-790-988-1	Sequence 1, Appli
C 45	40.6	1.6	2539	3	US-08-749-522-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-397-787-62
; Sequence 62, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; CURRENT FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-397-787-62

Query Match					9.7%; Score 240; DB 4; Length 240;
Best Local Similarity					100.0%; Pred. No. 7.8e-49;
Matches 240; Conservative					0; Mismatches 0; Indels 0; Gaps 0;
QY	1393	GTGTTGCCCTGTGATTAGTTCTGCTTTTAAACCCACTCCCTGGATGCAATTTTCCCTCC	1452		
Db	1	GTGTTGCCCTGTGATTAGTTCTGCTTTTAAACCCACTCCCTGGATGCAATTTTCCCTCC	60		
QY	1453	TTGCATTTCCCTTTTCTGAGTTCTATAGAGATCTGCATATGTTTTTCCCTTT	1512		
Db	61	TTGCATTTCCCTTTTCTGAGTTCTATAGAGATCTGCATATGTTTTTCCCTTT	120		
QY	1513	TTGCTTTGAGTGAAGTTTAAATAATCCACTCTGCTATTCACCTCTCGAACATC	1572		
Db	121	TTGCTTTGAGTGAAGTTTAAATAATCCACTCTGCTATTCACCTCTCGAACATC	180		
QY	1573	CCAAGCTGATCCCTGGCCTCTTTTCTCAGACTATGTTTCTTACCTGGACCTGAAT	1632		
Db	181	CCAAGCTGATCCCTGGCCTCTTTTCTCAGACTATGTTTCTTACCTGGACCTGAAT	240		

RESULT 2
US-09-566-921-125
; Sequence 125, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.

```

; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 125
; LENGTH: 2494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6682888 431588.3
; LOCATION: 2018-2291, 2487
; OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-125

Query Match      7.9%; Score 196.8; DB 4; Length 2494;
Best Local Similarity 63.1%; Pred. No. 4.6e-38;
Matches 303; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 815 TTAATATCTGCTTATGTTCTACCTTACAGTTTAAATCCCATGGATCAATCAATTAAT 874
DB 972 TTAAGAGTTGTATGCCAGACAGATATCAGTTTAAATCCCGTAAACCAATACACCTGAG 1031
QY 875 CATCATGACTACATTTGATTCCTCCCATCGCTGAAGGACAGAAATTCATTGTGTGGCATTTGTA 934
DB 1032 CATTTACTTTTATCACCTCTCCATCTCTGAAGGACAGGATTCACCTGTGTGGCTTATGTC 1091
QY 935 TTGTATGCCAGCTTATTCATTAATCTCTCTCAGATGATAGTAAGATCAAAAGATTT 994
DB 1092 TTAGACATCAACTCTATTGACATCTCTACTCTTAAATGTTGGCAAAAGTGAAGCAAGTT 1151
QY 995 CGAAGCGAGTTGTAAACGCTGTGTGTACATGTGGCTTCTCCTCAGATGATAGTAAGATCAAAAGATTT 1054
DB 1152 CACAAAGAGTATTAACCTGTGTATAGCATATGTGGCTTCTTACTTAAAGTGTATGAT 1211
QY 1055 ATGGATTTGATACAAAGGTGACCTTATAGAAATAGAGATGTGAGCCTGTGAGTCC 1114
DB 1212 TGCAGTGTGAGTCTTCAACAGCAACTTTTAAACATGAGTATGATCTATGACTTCTCAAAAGC 1271
QY 1115 AAGCTAGAGGAAGTCCAAAGAAACTTGGATTTGCTTTCTGACATCTCGGTGGTATGC 1174
DB 1272 CGGTCATGAATGCTCATTAATGCTAGGATTCCTATTTCCAAATATTTGATGTTGA 1331
QY 1175 AATTATCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAATCTTCTGCTGTG 1234
DB 1332 AATTATGCTTCAGATTTGGAATGGACCTGGACCCCATGAGGATATTCATCTCTCTGCACTG 1391
QY 1235 AGACGAATGCTATGGCTGCAGATGACTTCTTAGAGGATTTGCTTTTGGACAAATAGT 1294
DB 1392 AGCAGATGCTGGGCTGCAGATGATTTTAAAGATTTGCTTCTTGGAGAACTGTT 1451

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match      3.6%; Score 90.6; DB 1; Length 7218;
Best Local Similarity 14.4%; Pred. No. 2.8e-12;
Matches 81; Conservative 249; Mismatches 233; Indels 0; Gaps 0;

QY 1374 GCTCTGCTCTTTTGTGGGTTGTCCTGCTGATAGTATCTCTCTTTTAAACCCACTCCC 1433
DB 1064 GATVYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1123
QY 1434 TGAATGATTTTCCCTCTCTGATTTCCCTCTTTTCCTGGAGTTCTACTAGATCT 1493
DB 1124 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1183
QY 1494 GCATATGTTTCCCTCTTTGCTCTGAGATGAAGTTTAAATAATCAACCTCTGTCA 1553
DB 1184 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1243
QY 1554 TTCCACTCTCTGAACATCCCAAGCTGATTCCTGGCTCTTTTCTCAGACTATGTTCT 1513
DB 1244 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1303
QY 1614 TTACTTGGGACCTAGAACTGGATGGATGGCATGCTCTGATCAGATCAGACCTTTGA 1673
DB 1304 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1363
QY 1674 TTAATTCCTCTCTTAGGACCTTACCTCTGCTCTCTTTTGAATGCTCTTTTGTGTT 1733
DB 1364 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1423
QY 1734 CTCTCTCTCTCTAGTCCCTCTCTGATGAGTATGCTCTGCTAGTATGATGCT 1793
DB 1424 YYYYYYYYYYGTACCAATCTCTCTCTTAACTACTCTGATAGATAGTATGAT 1483
QY 1794 TTTTATGTAATGGCCACCGCATTTAGTATTAATAAATTTCTTTTAACTCTGTCAT 1853
DB 1484 AGTCATGCTACATCGCGTTTGTGAACTGAATAGATGCTCTAGAACGATGCTACGC 1543
QY 1854 AGTACATGCTGCTCTGTTCCATTTAGAGATTTGACAGAGGTTTTCAGTTTAGTATCTCA 1913
DB 1544 TAGTCACAAATCACCACTTTTATATTTAGATATATGTAATAAATAATAGTATGATTC 1603
QY 1914 ATCTTATTTAGTCTTGGGAAA 1936
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-49

Query Match      1.9%; Score 48; DB 4; Length 6306;
Best Local Similarity 49.2%; Pred. No. 0.05;
Matches 126; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY      5 ATCTTGTTATTAACCAACAGTTCACATAATTAATAATTTAGGAATTGAATTGTT 64
Db      450 AATTTAACTTTTAACTTAAATTTATCTCTAATACCTTATTTTATTACTATTATAAT 391

QY      65 AAGTTAAATTTGGTTTATATTTATGTTAGCAATTTATGTTGGTTCAAGATCAAACTACA 124
Db      390 AATATCATTTTTTAAATTAATAATTTTAACTAATTTTAAATATAATAAATACCATTT 331

QY      125 AATAATGTTATAGTCAAGAATCTATCTCTCTGCCCCCTTCAATAAATTTCTCCC 184
Db      330 TAAAAATTTTCTATCAATTTTTTTTAAAAATTTTCTTATTCATATTTAAATAAATTAATTT 271

QY      185 CTCTTCCCATTAAGTAAACCATATAAAATTTATATTTTACTTGGCTTTTAAAAATGTAACA 244
Db      270 TATTTCTTAAAAATCTTACATATTTTATTTATCTTCTTACCTTTTACCTAAATATAA 211

QY      245 AAGTACATATAAATTT 260
Db      210 CCTTCAATACCATTTT 195

RESULT 6
US-09-801-861-3
; Sequence 3, Application US/09801861
; Patent No. 6492154
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THERSOF
; FILE REFERENCE: CL001098
; CURRENT APPLICATION NUMBER: US/09/801,861
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53332
; TYPE: DNA
; ORGANISM: Human
US-09-801-861-3

Query Match      1.9%; Score 47.8; DB 4; Length 53332;
Best Local Similarity 44.9%; Pred. No. 0.12;
Matches 181; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY      6 TTCTTGTTATTAACCAACAGTTCACATAATTAATAATTTAGGAATTGAATTGTTA 65
Db      31258 TTGCTTTATCATCTACATATATTTTAAATAAATTTCTAAAAATGCGCAATCACGTATATA 31317

QY      66 AGTTAAATTTGGTTTATATTTATGTTAGCAATTTATGTTGGTTCAAGATCAAACTACA 125
Db      31318 TTTTATATATATTTTATATTTTATATATTTTATATATTTTATATATTTTATATATTTATAT 31377

QY      126 AATAATGTTATAGTCAAGAATCTATCTCTCTGCCCCCTTCAATAAATTTCTCCC 185
Db      31378 TTTATATATTTTATATTTTATATATTTTATATATTTTATATATTTTATATATTTTATATA 31437

QY      186 TCTTCCCATTAAGTAAACCATATAAAATTTTATATTTTACTTGGCTTTTAAAAATGTAACA 245
Db      31438 TATTTATATATTTTATATTTTATATTTTATATATTTTATATATTTTATATATTTTATATA 31497

QY      246 AGTACATATAAATTTGCTGCTACTCCCTTCTTCTAGAGAGTGGTGAAGAAACATGTTATAT 305
Db      31498 TATATATATATATATTTTATATTTTATATATTTTATATATTTTATATATTTTATATATAT 31557

; ORGANISM: Homo sapiens
; FEATURE: misc_feature
; LOCATION: (1)..(551)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-122

Query Match      2.2%; Score 53.6; DB 3; Length 551;
Best Local Similarity 80.0%; Pred. No. 0.00098;
Matches 96; Conservative 0; Mismatches 20; Indels 4; Gaps 3;

QY      845 TTTAATCCCATGGAATC-AATCAAAATTAATCAATCATGACTACATTCATTCATCCCATCGCT 903
Db      427 TTAATCCCATGGAATCAATCAAAATTAATCAATCATGACTACCTTGGTTCCTCATCGGT 486

QY      904 GAA--GGACAGAAATCATGTGTGCATTTGTATTTGAT--GCCAGCTCTATTCATTAATCTT 960
Db      487 TGAAGGGAGCAGNATTCATTTGGGGGCGCATTTGATTTGATNNCAGNTTTTATTCATTAATCTT 546

RESULT 5
US-10-204-708-49/C
; Sequence 49, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 49
; LENGTH: 6306
; TYPE: DNA
```


US-08-107-755A-32/c
; Sequence 32, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: NO. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-107-755A-32
Query Match 1.8%; Score 45.2; DB 1; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
QY 9 TTGTATTATAACCAACAGTCTACTAATTAATAATTTAGGAATTTGTAAGT 68
DB 436 TAGGAAAACATTATTAAAGTTACCAAAATTTATTTTAGATATTATTAAATTTTAAAT 377
QY 69 TAATTTGGTTTATATTATCTTTAG-----CATTATGTGTTTCAAGATCAAACTA 122
DB 376 TTACTAAATATTATAAAATTTAGTCATTATTATACATGATTCACATTTAAAATTT 317
QY 123 CAAATAATGTATAGTCAAGAACTATCTTCCTCTCTGCCCTTCAAAATAAATTTCTC 182
DB 316 CTATAGAATGGTAGTATAATATTACATTATGCTATTTTGTATAGATATATCTA 257
QY 183 CCCTCTCCCATAGTACCAATATAAATTTATTTTACTTGGCTTTTAAATATCTAA 242
DB 256 AATATGTTATATTTTAAATTTTGTATATAAATTTAAATTAATTAATTTAAATTTGAAA 197
QY 243 CAAAGTACATATAAATTTGGCTGCTACTCCCTCTTAGAGAAAGTGGTAGAAAACCTATGTTA 302
DB 196 TATATAAACTTTTATATTTTCTGGAATATTATTATAAATATTATATCATATAATATTA 137
QY 303 TATTGACTTATCAGACATTTGTTTAACTGACATGGCAATTTTCTGCTACAAATGTTCCAGC 362

DB 136 TATGAANTCTCTCTAAATAACTAATTTTAAATATATATTAATAATAACATTATCTC 77
QY 363 AGTTAATAATCTTTCATATATATCATTTTGCATTTTGTGTCAGTATATCA 410
DB 76 TGTATTATTATTATTTTAAATTTATATAATTTTAAATATTTTAAATTTTAAAT 29
RESULT 10
US-08-544-332-32/c
; Sequence 32, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: NO. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UF114.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-544-332-32
Query Match 1.8%; Score 45.2; DB 2; Length 660;
Best Local Similarity 46.3%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
QY 9 TTGTATTATAACCAACAGTCTACTAATTAATAATTTAGGAATTTGTAAGT 68
DB 436 TAGGAAAACATTATTAAAGTTACCAAAATTTATTTTAGATATTATTAAATTTTAAAT 377
QY 69 TAATTTGGTTTATATTATCTTTAG-----CATTATGTGTTTCAAGATCAAACTA 122

US-07-991-867B-8
LOCATION: 852...1511
Query Match
Best Local Similarity 1.8%; Score 45.2; DB 1; Length 1511;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
Y 9 TTGTATTATTAACCAATACAGTTCACATTAATAAATTAAGTAAATTTAGGAATGGAATGTTAAAGT 68
Db 1287 TAGGAAAACATTATTAAAGTTACCAAAATTTATTTTAGATATTATTAATTTTAAAT 1228
Y 69 TAATTTGGTTTATTTATTTAGTTAG-----CAATTTATGTGTTTCAAGATCAAAATCTA 122
Db 1227 TTTCTAAATTTATTAATAAATTTATAGTCATTTATATACATTTACAAATTTAAAT 1168
Y 123 CAAATATAGTATAGTCAAGATCTATCTCTCTCTGCGCCCTTCAATTAATTTCTC 182
Db 1167 CTATAGAAATGTTAGTATTAATTTATTTATTTATTTTGTATTAAGATATATCTA 1108
Y 183 CCCTCTTTCCCATTAAGTAAACCAATATAAATTTATTTTACTTCTGCGCTTTTAAATATGTA 242
Db 1107 AATATGTTATTTTATTTTATTTTATTAATAAATTTAAATTAATAATTTAAATTTGAA 1048
Y 243 CAAAGTACATATAAATTTGCTGCTACTCCCTTTAGAGAAGTGGTAGAAAATCTATGTTA 302
Db 1047 TATATAAACTTTTAAATTTTCTGGAATATTTTAAATATTTATTTATCATATATATTA 988
Y 303 TATGACTTATCAGACATTTGTTTAACTGACATGGCAATTTTCTGCTACAAATGTTCCAGC 362
Db 987 TATGCAATTTCTTAAATTTAACTAAATTTTAAATATTTTAAATATTTTAAATTTTAA 928
Y 363 AGTTAATAATCTTTGCATATATCATTTTGCATTTTGTGTCAGTATATCA 410
Db 927 TGTATTATTTATTTTAAATTAATAATTTTAAATATTTTAAATATTTTAAATTTTAA 880
RESULT 13
US-08-107-755A-8/c
Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C2
TELECOMMUNICATION INFORMATION:

TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18...218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234...782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852...1511
US-08-107-755A-8
Query Match
Best Local Similarity 1.8%; Score 45.2; DB 1; Length 1511;
Matches 189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
Y 9 TTGTATTATTAACCAATACAGTTCACATTAATAAATTTAGGAATGGAATGTTAAAGT 68
Db 1287 TAGGAAAACATTATTAAAGTTACCAAAATTTATTTTAGATATTATTAATTTTAAAT 1228
Y 69 TAATTTGGTTTATTTATTTAGTTAG-----CAATTTATGTGTTTCAAGATCAAAATCTA 122
Db 1227 TTTCTAAATTTATTAATAAATTTATAGTCATTTATATACATTTACAAATTTAAAT 1168
Y 123 CAAATATAGTATAGTCAAGATCTATCTCTCTCTGCGCCCTTCAATTAATTTCTC 182
Db 1167 CTATAGAAATGTTAGTATTAATTTATTTATTTTGTATTAAGATATATCTA 1108
Y 183 CCCTCTTTCCCATTAAGTAAACCAATATAAATTTATTTTACTTCTGCTACAAATGTTCCAGC 242
Db 1107 AATATGTTATTTTATTTTATTTTATTAATAAATTTAAATTAATAATTTAAATTTGAA 1048
Y 243 CAAAGTACATATAAATTTGCTGCTACTCCCTTTAGAGAAGTGGTAGAAAATCTATGTTA 302
Db 1047 TATATAAACTTTTAAATTTTCTGGAATATTTTAAATATTTATTTATCATATATATTA 988
Y 303 TATGACTTATCAGACATTTGTTTAACTGACATGGCAATTTTCTGCTACAAATGTTCCAGC 362
Db 987 TATGCAATTTCTTAAATTTAACTAAATTTTAAATATTTTAAATATTTTAAATTTTAA 928
Y 363 AGTTAATAATCTTTGCATATATCATTTTGCATTTTGTGTCAGTATATCA 410
Db 927 TGTATTATTTATTTTAAATTAATAATTTTAAATATTTTAAATATTTTAAATTTTAA 880
RESULT 14
US-08-544-332-8/c
Sequence 8, Application US/08544332
Patent No. 5935777
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard H. Bencen
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/544,332
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,867
FILING DATE: 07-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,755
FILING DATE: 19-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,695
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,594
FILING DATE: 13-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Geirard H.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: UF114.C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-544-332-8

```

Query Match	1.8%; Score 45.2; DB 2; Length 1511;
Best Local Similarity	46.3%; Pred. No. 0.15;
Matches 189; Conservative	0; Mismatches 213; Indels 6; Gaps 1;
QY	9 TTGTATTATAACCAATAACAGTTCCTACTAAATAAAATAAAATTTAGGAAATGGAAATGTTAAAGT 68
Db	1287 TAGGAAAAACATATTAAAGTTACCAAAATTAATTTTATAGATATATTAAATTTTAAAT 1228
QY	69 TAAATTTGGTTTTATATTATGTTTAG-----CATTTATGTGGTTCAGAAATCAAAATCTA 122
Db	1227 TTACTAAATTTATTAATAAAATATTAGTCATTATATTACATGATTCAAATTTAAAAAAT 1168
QY	123 CAAAATAATGTATAGTCAAAGAAATCTATCTTCTCTGCCCTTCGAAATAAAATTTCTC 182
Db	1167 CTATAGAATGTGGTAGTATAATAATTACTATATTGCTATTTTGTATAAGATATACTA 1108
QY	183 CCCTCTCCCATAGTAGTAACCATATAAAATTTATATTTTACTTGCGCTTTAAATATGTAA 242
Db	1107 AATATGTTATATTTTTTAAATTTTGGTTAAANAATTTAAATTAATATTTTAAATTTGAAA 1048
QY	243 CAAAGTACATAAAATTTGCTGCTACTCCCTTCTTAGAGAAGTGGTAGAAAACCTATGTTA 302
Db	1047 TATATAAAGCTTTTAATATTTTCTGGAATATATTATAAAATATTATTCATATATATATTA 988

Qy	303	TATTGACCTATCAGACATGTTTAACTGCACATGCGCATTTTCTGCTACAAATGTTT
Db	987	TATGCAATCTCTAAATTAACATAATTTTTTAAATATTAATTAATTAACATTT
Qy	363	AGTTAATAATCTTGCATATATCATTTTGGCATTTTTTCTCAGTATATCA 410
Db	927	TGTTTATATTATTTTTTTTAAATATAATATTTTAAATATTTATTA 880

RESULT 15
 US-09-370-861A-8/c
 ; Sequence 8, Application US/09370861A
 ; Patent No. 6410221
 ; GENERAL INFORMATION:
 ; APPLICANT: Moyer, Richard W.
 ; APPLICANT: Hall, Richard L.
 ; APPLICANT: Gruidl, Michael E.
 ; TITLE OF INVENTION: No. 6410221el Entomopoxvirus Expression System
 ; FILE REFERENCE: UF114.C4.D1
 ; CURRENT APPLICATION NUMBER: US/09/370,861A
 ; CURRENT FILING DATE: 1993-08-09
 ; PRIOR APPLICATION NUMBER: US 07/991,867
 ; PRIOR FILING DATE: 1992-12-07
 ; PRIOR APPLICATION NUMBER: US 08/107,755
 ; PRIOR FILING DATE: 1993-08-19
 ; PRIOR APPLICATION NUMBER: WO 92/14818
 ; PRIOR FILING DATE: 1992-02-12
 ; PRIOR APPLICATION NUMBER: US 07/827,685
 ; PRIOR FILING DATE: 1992-01-30
 ; PRIOR APPLICATION NUMBER: US 07/657,584
 ; PRIOR FILING DATE: 1991-02-19
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 1511
 ; TYPE: DNA
 ; ORGANISM: Amsacta moorei entomopoxvirus
 US-09-370-861A-8

Query Match	1.8%; Score 45.2; DB 4; Length 1511;
Best Local Similarity	46.3%; Pred. No. 0.15;
Matches	189; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
QY	9 TTGTATTATAACCATACAGCTTCACATAATTAATTAATTTAGGAATTGAATTTGTTAAGT 68
Db	1287 TAGGAAAAACATTTATTAAAGTTACCAATTTATTTTGTAGATATTATTAATTTTITTAAT 1228
QY	69 TAAATTTGGTTTTATATTATGTTTAG-----CATTTATGTGGTTCAAAGATCAAATCTA 122
Db	1227 TTACTAAATTTATTAATAAAATTTATGTCAATTTATATTACATGATTCACAATTTAAAAAATT 1168
QY	123 CAAAAATATGTATAGTCAAAGATCTATCTTCCTCTCTCGCCCTTCAAATAAATTTCTC 182
Db	1167 CTATAGAATGTGGTAGTATAATATTACTATATTTGCTATTTTGTGTATAGATATATCTA 1108
QY	183 CCCTCTTCCCATTAGTAGTAACCATATAAATTTATATTTTACTTCGCTTTTAAAAATATGTAA 242
Db	1107 AATATGTTATATTTTITTAATTTTGTGTATAAAATTTAAATTAATAATTTTAAATTTGAAA 1048
QY	243 CAAAGTACATATAAAATTTGGTGTCTACCTCCCTTCTTAGAGAAAGTGGTAGAAAACATATGTTA 302
Db	1047 TATATAAACCTTTTAATATTTTCTGGAAATATTTATTTAAATATTATTATTCATAATATATTA 988
QY	303 TATTGACCTTATCAGACATCTGTTTAACTGACATGGCATTTTTTCTGCTACAATGTTCACGC 362
Db	987 TATGCAATTCCTCTAAATTAACATAATTTTTTAAATATATATTAATTAATTAACATTAATCTC 928
QY	363 AGTTAAATAATCTTTGCAATATATCATTTTGGCATTTTTCGTCAGTATATCA 410
Db	927 TGTATTATATATTTTTTTTTAAATTAATAATTTTAAAAATATTTATTTA 880

Search completed: April 26, 2004, 02:11:16
Job time : 125.228 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:47:42 ; Search time 660.01 seconds

15988.422 Million cell updates/sec
(without alignments)

Title: US-10-051-835-18

Perfect score: 2484
Sequence: 1 gtttatcttctgtattataac.....ataaaattctctataaaaactc 2484

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : N_Geneseq_29Jan04:
1:   geneseq1980s.*
2:   geneseq1990s.*
3:   geneseq2000s.*
4:   geneseq2001as.*
5:   geneseq2001bs.*
6:   geneseq2002s.*
7:   geneseq2003as.*
8:   geneseq2003bs.*
9:   geneseq2003cs.*
10:  geneseq2004s.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2484	100.0	2484	7	ABx63692	Human cDN
2	1358	54.7	2741	8	ACd42248	Human hpn
3	1358	54.7	2741	9	ADs25678	Human cDN
4	1335.2	49.7	2227	4	ABs60823	Human can
5	456.6	18.4	2068	7	ABx63519	Human can
6	455	18.3	1335	2	ABx633941	Human cDN
7	455	18.3	1335	2	ABx633941	Human HCM
8	448.6	18.1	1332	1	ABn59995	Novel hum
9	415.4	16.7	567	4	AAu81259	Sequence
10	368	14.8	482	6	AASe6047	Human can
11	354	14.3	2515	7	Abj180943	Human ova
12	331.2	13.3	580	4	ABx63520	Human can
13	315.4	12.7	977	4	AAh34468	Human col
14	311	12.5	3276	4	AASe27231	cDNA enco
15	311	12.5	3276	9	ADb93409	Human cDN
16	281	11.3	600	7	ABx63603	Human cDN
17	272	11.0	272	6	ABk29513	Colon ade
18	240	9.7	240	6	ABT06900	Human ova
19	240	9.7	240	7	ABx72778	Human ova
20	196.8	7.9	2058	6	ABs76439	cDNA enco
21	196.8	7.9	2058	9	ADD18927	Human dis
22	196.8	7.9	2062	3	AAa08592	Human cyt
23	196.8	7.9	2062	7	ABx63510	Human cDN

ALIGNMENTS

RESULT 1

ABX63692
ID ABX63692 standard: cDNA: 2484 bp.

AC ABX63692:

DT 26-FEB-2003 (first entry)

Human cDNA #692 differentially expressed in activated vascular tissue.

Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
hypertensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
gene therapy; vascular disease; cancer; coronary; artery disease;
hypertension; diabetes; pre-eclampsia; restenosis;
ischaemia-reperfusion injury; stroke.

OS Homo sapiens.

AA
PN
US2002137081-A1.

26-SEP-2002

08-VAN-2002: 2002US-00044090-XX

XX
PR 28-JUL-2000: 2000US-0222469P.

PR 08-JAN-2001; 2001US-0260483P.
yy

FA (BAND 7) BANDMAN O.
XX

PI Bandman O;
XX

DR WPL; 2003-110597/10.
XX

Combination for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, comprises several cDNAs that are differentially expressed in activated vascular tissue

XX
DS
Claim 1. Date: 1800. English

XX This invention relates to a combination comprising several cDNAs that are
CC differentially expressed in activated vascular tissue. The invention also
CC discloses a high throughput method for detecting differentially expressed
CC cDNAs in a sample. The cDNAs of the invention may have
CC antiarteriosclerotic; cytostatic; cardiac; hypertensive; antidiabetic;
CC antineoplastic; and/or neuroprotective activities and may be
CC used in the treatment of atherosclerosis, hypertension, diabetes, and
CC cancer.

used in gene therapy. The cDNAs of the invention may be used in a high-throughput methods for detecting differential expression of one or more cDNAs in a sample, or screening several molecules or compounds to identify a molecule or compound that specifically binds a cDNA of the invention. A protein encoded by the cDNA may be used to screen several molecules or compounds to identify a ligand that specifically binds to the protein, or to produce or purify an antibody to the protein that can be used to detect a protein in a sample or purify a natural or recombinant protein from a sample. The nucleotides may be useful for diagnosing, staging, treating, or monitoring the progression of treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion injury, retinosis, or stroke. The cDNAs can also be used for large-scale genetic or gene expression analysis of several new nucleic acid molecules. Antibodies to the proteins encoded by the cDNAs are useful for diagnosing pre-pathologic disorders, and chronic or acute diseases associated with abnormalities in the expression, amount or distribution of the protein. The present sequence represents a cDNA of the invention that is differentially expressed in activated vascular tissue. Note: The sequence data for this patent did not form part of the specification, but was obtained in electronic format directly from USPTO at <http://seqdata.uspto.gov/sequence.html?DocID=20020137081>

XX
Sequence 2184 BP: 657 A; 488 C; 436 G; 903 T; 0 U; 0 Other;

Query Match	100.0%;	Score 2484;	DB 7;	Length 2484;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2484;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	GTTTATCTCTGTPATTATATAACCATAAACAGTTCACATAATTAATTAATTAAGGAATTGAAT	60
Db	1	GTTTATCTCTGTTATTATTAACCATAAACAGTTCACATAATTAATTAATTAAGGAATTGAAT	60
Qy	61	TGTTAAGTTAAATTTGGTTTTATTTATTTATGTTTAGCATTTATGTGGTTCAAAGATCAAATC	120
Db	61	TGTTAAGTTAAATTTGGTTTTATTTATTTATGTTTAGCATTTATGTGGTTCAAAGATCAAATC	120
Qy	121	TACAAATAATGCTATAGTCAAGATCTATCTTCCTTCTCTGGCCCTTCAAAATAAAATTC	180
Db	121	TCAAAATAATGCTATAGTCAAGAAATCTATCTTCCTTCTCTGGCCCTTCAAAATAAAATTC	180
Qy	181	TCCCTCTCTTCCCATAGTAACCATATAAAATTTATATTTTACTTTCGCTTTTAAAAATATGT	240
Db	181	TCCCTCTCTTCCCATAGTAACCATATAAAATTTATATTTTACTTTCGCTTTTAAAAATATGT	240
Qy	241	AACAAAGTACATATAAAATTTGCTGCTACTCCCTCTCTTAGAGAAGTGGTAGAAAACTATGT	300
Db	241	AACAAAGTACATATAAAATTTGCTGCTACTCCCTCTCTTAGAGAAGTGGTAGAAAACTATGT	300
Qy	301	TATATTGACTATPCAGACATGTTTAACTGACATGGCATTTTTTCGTACAATAATGTTC	360
Db	301	TATATTGACTATPCAGACATGTTTAACTGACATGGCATTTTTTCGTACAATAATGTTC	360
Qy	361	GCAGTTAAATAATCTTTTGCATATATCATTTTGCATTTTGTTCAGTATATCAGTGGGACAGA	420
Db	361	GCAGTTAAATAATCTTTTGCATATATCATTTTGCATTTTGTTCAGTATATCAGTGGGACAGA	420
Qy	421	TTCCGAACAGAGAAGTGTCTAGATGAAGAGTAAAGTCATCTCACCTTGGAGCCCTTCCTTC	480
Db	421	TTCCGAACAGAGAAGTGTCTAGATGAAGAGTAAAGTCATCTCACCTTGGAGCCCTTCCTTC	480
Qy	481	CTTGGCGCTCTGCTTTGATGGTACCAAGCTTTTTTTTCTGAAATGCTCTTACTTCCATCTGTGG	540
Db	481	CTTGGCGCTCTGCTTTGATGGTACCAAGCTTTTTTTTCTGAAATGCTCTTACTTCCATCTGTGG	540
Qy	541	GCAGATTCGCTCTCTTATCTAGAACACTTCAGGCATATCCTTATCTCTTTAGCAGACT	600
Db	541	GCAGATTCGCTCTCTTATCTAGAACACTTCAGGCATATCCTTATCTCTTTAGCAGACT	600
Qy	601	CACCTCTGCTANTTTAGCTCTCAGATTAATCATGTCTCTTCTCCATCAGAGGCTTACTTGAC	660
Db	601	CACCTCTGCTANTTTAGCTCTCAGATTAATCATGTCTCTTCTCCATCAGAGGCTTACTTGAC	660

Db 1964 TTGACTTGCTTTTGTGTTCTTCTTCATCTTAGTCCCTCTTCANGCAGTAGTGCAT 2023
Qy 1775 GCTAGGTAGAGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTACATAACTTT 1834
Db 2024 GCTAGGTAGAGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTACATAACTTT 2083
Qy 1835 CTTTAAACATCTGTCATAGTACATGCTGCTCTGTTCCATTTAGAGATTGACAGAGT 1894
Db 2084 CTTTAAACATCTGTCATAGTACATGCTGCTCTGTTCCATTTAGAGATTGACAGAGT 2143
Qy 1895 TTCAGTTTATGATACATCAATCTTATTTTAGTGTCTGGGAAATCAATTGAGAAATATCACA 1954
Db 2144 TTCAGTTTATGATACATCAATCTTATTTTAGTGTCTGGGAAATCAATTGAGAAATATCACA 2203
Qy 1955 TCCTCCCAATCTCTTACTCAAAATGCTGGGAAACTCTCATGTTACTACTTTGTTG 2014
Db 2204 TCCTCCCAATCTCTTACTCAAAATGCTGGGAAACTCTCATGTTACTACTTTGTTG 2263
Qy 2015 CTCCTACTGCTGCTGTTGTTTCCCATCCCTCTCTTCCCTCATGTTAGTGTCTCTCT 2074
Db 2264 CTCCTACTGCTGCTGTTGTTTCCCATCCCTCTCTTCCCTCATGTTAGTGTCTCTCT 2323
Qy 2075 AATATTAGGCTTGGTTCAGATTTTCAGTGTCCAAATATTCCTCTTCCCTCTGTTGCCCTT 2134
Db 2324 AATATTAGGCTTGGTTCAGATTTTCAGTGTCCAAATATTCCTCTTCCCTCTGTTGCCCTT 2383
Qy 2135 TCCTGAGATAATCACTAAGATATTTTGTGTTCTTTCTCAGGGAATCTAAGGAGGA 2194
Db 2384 TCCTGAGATAATCACTAAGATATTTTGTGTTCTTTCTCAGGGAATCTAAGGAGGA 2443
Qy 2195 AATTATCACTGTGCAAGGAAAAAATAGATATGTGAAGGTTCACTGTAATTTCTCTC 2254
Db 2444 AATTATCACTGTGCAAGGAAAAAATAGATATGTGAAGGTTCACTGTAATTTCTCTC 2503
Qy 2255 ACATCAGAGATTAATTAATTCAGAGGAGGAAACACAGACCAAGAGAGTATCTAAG 2314
Db 2504 ACATCAGAGATTAATTAATTCAGAGGAGGAAACACAGACCAAGAGAGTATCTAAG 2563
Qy 2315 ACCAAGGGATGTGTTTATTAATGCTAGAGTGAAGAAATGATAGACATTTGATGAC 2374
Db 2564 ACCAAGGGATGTGTTTATTAATGCTAGAGTGAAGAAATGATAGACATTTGATGAC 2623
Qy 2375 TTGTAATACTAGATAATACATGATTTAGTCAATTTGTAATAATAATAATTTT 2434
Db 2624 TTGTAATACTAGATAATACATGATTTAGTCAATTTGTAATAATAATAATTTT 2683
Qy 2435 CTTGGATTATGTTCTGTATCTGTGAAAAAATAAATTTCTTATAAACTC 2484
Db 2684 CTTGGATTATGTTCTGTATCTGTGAAAAAATAAATTTCTTATAAACTC 2733

RESULT 3
ID ADE25678
XX ADE25678 standard; cDNA; 2741 BP.
XX AC ADE25678;
XX DT 29-JAN-2004 (first entry)
XX DE Human cDNA differentially expressed in foam cells #82.
XX KW Human; ss; differential expression; foam cell; LPS; lipopolysaccharide;
XX KW cardiovascular disease; atherosclerosis.
XX OS Homo sapiens.
XX PN US2003194721-A1.
XX PD 16-OCT-2003.
XX PF 18-SEP-2002; 2002US-00247671.
XX PR 19-SEP-2001; 2001US-0323784P.

XX (INCY-) INCYTE GENOMICS INC.
XX Mikita T, Shiffman D, Porter JG, Kaser MR;
XX WPI; 2003-875398/81.
XX Combination containing several polynucleotide that are differentially
XX expressed in foam cells and complements of the polynucleotides, useful
XX for diagnosing cardiovascular disease or atherosclerosis.
XX Claim 1; SEQ ID NO 82; 37pp; English.
XX The invention relates to a combination comprising several polynucleotides
XX having any one of 127 sequences (S1) such as the sequence of human
XX calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4
XX hydrolase, human C9orf142 protein mRNA, human K+ channel beat 2 subunit
XX mRNA, etc., and their complements. The cDNAs are differentially expressed
XX in LPS (lipopolysaccharide)-treated foam cells. Also included are
XX sequences, an expression vector containing the nucleic acids, a host cell
XX containing the vector, a purified polypeptide appearing as ADE25750 and
XX ADE25751, producing a protein by culturing the host cell, and a
XX composition comprising a purified antibody that specifically binds to the
XX polynucleotides in a sample. The sample is from a subject with
XX atherosclerosis and comparison with a standard defines early, mid or late
XX stages of the disorder. The foam cell-expressed nucleic acids are useful
XX for high throughput screening of a library of molecules or compounds to
XX identify a ligand which binds a polynucleotide. The library is chosen
XX from DNA molecules, peptides, proteins and RNA molecules. The protein is
XX useful for a high throughput screening of library of molecules or
XX compounds to identify at least one ligand which specifically binds a
XX protein, for purifying a ligand from a sample for making an antibody. The
XX foam cell-expressed nucleic acids are useful for diagnosing
XX cardiovascular disorder. The foam cell-expressed nucleic acids are useful
XX as elements on a microarray which can be used for detecting related
XX polynucleotide in a sample, diagnosing cardiovascular disease,
XX atherosclerosis. The present sequence represents a cDNA whose expression
XX is upregulated in LPS treated foam cells.
XX Sequence 2741 BP; 841 A; 452 C; 530 G; 918 T; 0 U; 0 Other;
Qy Query Match 54.7%; Score 1358; DB 9; Length 2741;
Best Local Similarity 99.9%; Pred. No. 5.6e-296;
Matches 1369; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1116 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTGCTCTTCTGACATCTCGTGGTTAGCA 1175
Db 1364 AGCTAGAGGAAGTCCAAAGAAAACTTGGATTGCTCTTCTGACATCTCGTGGTTAGCA 1423
Qy 1176 ATTATTCCTCTGAGTGGGAGCTGGACCTGTAAAGGATGTTCTAATTTCTTCTCTCTGA 1235
Db 1424 ATTATTCCTCTGAGTGGGAGCTGGACCTGTAAAGGATGTTCTAATTTCTTCTCTCTGA 1483
Qy 1236 GACGAATGCTATGGGCTCCAGATGACATTTCTTAGAGGATTTGCCCTTTTGAGCAATAGGTA 1295
Db 1484 GACGAATGCTATGGGCTCCAGATGACATTTCTTAGAGGATTTGCCCTTTTGAGCAATAGGTA 1543
Qy 1296 GATG-CTTTGTTGGTGTGGAAGCTTGGAGCGTTCAGGTAGTTCGCTACTTTCGTTCTGG 1354
Db 1544 GATGTTTGGTGTGGAAGCTTGGAGCGTTCAGGTAGTTCGCTACTTTCGTTCTGG 1603
Qy 1355 ATCTATTAATAACCTGGCAGCTCTCTGCTTTTGTGGTGTGTTGCCCTGTGATAGTTTC 1414
Db 1604 ATCTATTAATAACCTGGCAGCTCTCTGCTTTTGTGGTGTGTTGCCCTGTGATAGTTTC 1663
Qy 1415 TCGTTTTTAACCCACTCCCTGAGTGCATTTTCCCTCCCTGCTTCATTTCCCTCTGG 1474
Db 1664 TCGTTTTTAACCCACTCCCTGAGTGCATTTTCCCTCCCTGCTTCATTTCCCTCTGG 1723
Qy 1475 AGTTCATACCTAGAGAACTGCACTATGTTTTCCTTTTTCCTTTGCTTTGAGATGAAGTTT 1534

Db 1724 AGTTCATACAGAGAACTGACACTATGTTTTTCCCTTTTGTCTTGAGATGAAAGTTTGA 1783
Qy 1535 AAAATAATCACTCTGTCATTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCTCT 1594
Db 1784 AAATAATCACTCTGTCATTTCCACTCTCTGAACATCCCAAGCTGTATCCCTGGCTCT 1843
Qy 1595 TTCTCAGACTATGTTCTTTACTTTGGACCTAGAACCTGGAATGGATTTGGATTTGCTCCT 1654
Db 1844 TTCTCAGACTATGTTCTTTACTTTGGACCTAGAACCTGGAATGGATTTGGATTTGCTCCT 1903
Qy 1655 GATCAGATGAGACCTTTCATATTTTGGCCCTTCCCTTAGGACCTTACACTCTGCTTTCT 1714
Db 1904 GATCAGATGAGACCTTTCATATTTTGGCCCTTCCCTTAGGACCTTACACTCTGCTTTCT 1963
Qy 1715 TTGACTTGCCTTTTGTGTTCTTTCTTCACTTAGTCCCTCTTCAATGAGTATGTCATT 1774
Db 1964 TTGACTTGCCTTTTGTGTTCTTTCTTCACTTAGTCCCTCTTCAATGAGTATGTCATT 2023
Qy 1775 GCTAGGTAGAGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTAGATAAATCTT 1834
Db 2024 GCTAGGTAGAGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTAGATAAATCTT 2083
Qy 1835 CTTTTAACAACTGTGTCATAGTACATGCTGCTCTGTTCCATTTAGAGATTTGACAGAGT 1894
Db 2084 CTTTTAACAACTGTGTCATAGTACATGCTGCTCTGTTCCATTTAGAGATTTGACAGAGT 2143
Qy 1895 TTCAGTTAGTATACCAATCTTATTTAGTGTCTGTTGGAAATCAATTCAGAAATATCACA 1954
Db 2144 TTCAGTTAGTATACCAATCTTATTTAGTGTCTGTTGGAAATCAATTCAGAAATATCACA 2203
Qy 1955 TCCTCTCAATTCCTCTTACTCAATTTGCTGGAAACTCTCATGTTACTTAACCTTTGTTG 2014
Db 2204 TCCTCTCAATTCCTCTTACTCAATTTGCTGGAAACTCTCATGTTACTTAACCTTTGTTG 2263
Qy 2015 CTCCTAATCTGCACTCTGTTTCCCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2074
Db 2264 CTCCTAATCTGCACTCTGTTTCCCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2323
Qy 2075 AATATTAGCTTGGTTGAGATTTTCAGTGTGTCCTCAATTTCTCTCTCTCTCTCTCTCT 2134
Db 2324 AATATTAGCTTGGTTGAGATTTTCAGTGTGTCCTCAATTTCTCTCTCTCTCTCTCTCT 2383
Qy 2135 TCCTGAGTAATPCCACTAAGAAATATTTTGTGTTTCTTTCTCAGGGAATCTAAGGAGGA 2194
Db 2384 TCCTGAGTAATPCCACTAAGAAATATTTTGTGTTTCTTTCTCAGGGAATCTAAGGAGGA 2443
Qy 2195 AATTATCACTCTGCACAGGAAATATAGATATGTGAAGGTTTCAGTAAATTTCTCTC 2254
Db 2444 AATTATCACTCTGCACAGGAAATATAGATATGTGAAGGTTTCAGTAAATTTCTCTC 2503
Qy 2255 ACATCACAGAGATTAATAATTCAGAAAGGAGAAACACAGACCCAAAGAGATATCTAAG 2314
Db 2504 ACATCACAGAGATTAATAATTCAGAAAGGAGAAACACAGACCCAAAGAGATATCTAAG 2563
Qy 2315 ACCAAGGATGTTTATTAATCTTAGTATGAGGATGAAGATGTCATAGACATTTCTAGTAC 2374
Db 2564 ACCAAGGATGTTTATTAATCTTAGTATGAGGATGAAGATGTCATAGACATTTCTAGTAC 2623
Qy 2375 TTGTAATAACTAGAAATAACATGATTTTAGTCATATTTGTGAATAATAATAATAATTTT 2434
Db 2624 TTGTAATAACTAGAAATAACATGATTTTAGTCATATTTGTGAATAATAATAATAATTTT 2683
Qy 2435 CTTGGATTATGTTCTGTATCTGTGAAATAATAATTTCTTATAAATCTC 2484
Db 2684 CTTGGATTATGTTCTGTATCTGTGAAATAATAATTTCTTATAAATCTC 2733

RESULT 4
ID AAS60823
XX AAS60823 standard; cDNA; 2527 BP.
AC AAS60823;

XX 29-JAN-2002 (first entry)
XX Human cancer agent-sensitive marker #325.
XX Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
XX squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
XX lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
XX Hodgkin's disease; glioma; ss.
XX Homo sapiens.
XX WO200179556-A2.
XX 25-OCT-2001.
XX 13-APR-2001; 2001WO-US012132.
XX 14-APR-2000; 2000US-0197538P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lillie J, Brown JL, Bolt A, Van Huffel C;
XX WPI; 2001-602933/68.
XX Novel nucleic acid, used as a marker to determine the effectiveness of
XX using TAXOL to treat cancer cell growth in individuals.
XX Claim 1; Page 338-339; 527pp; English.
XX The invention relates to 1046 novel nucleic acids which are used as
XX markers for determining the sensitivity of a cancer cell to the
XX anticancer agent TAXOL. Cancer cells can be treated with TAXOL when they
XX are shown to express one of the 242 sensitivity markers of the cells are
XX shown not to express one of the 804 resistance markers. The methods can
XX be used to determine the effectiveness of TAXOL in the treatment of
XX cancer cell growth in an individual. The markers can be used as targets
XX in developing anti-cancer agents such as chemotherapeutic compounds. The
XX markers can also be used as targets in developing treatments for cancer,
XX particularly those cancers which display resistance to agents and exhibit
XX expression of the markers. The anticancer agents developed by the novel
XX method can be used to treat cancer. Probes based on the markers can be
XX used to detect transcripts or genomic sequences corresponding to the
XX markers, in the identification of cells or tissues which mis-express the
XX protein. Cancers which may be targeted include carcinoma (e.g. squamous
XX cell carcinoma), sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic
XX leukaemia), lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's
XX disease and tumours (e.g. glioma). The present sequence is one of the
XX 1046 novel cancer cell markers

XX SQ Sequence 2527 BP; 742 A; 431 C; 513 G; 824 T; 0 U; 17 Other;
Query Match 49.7%; Score 1235.2; DB 4; Length 2527;
Best Local Similarity 98.6%; Pred. No. 2.6e-268;
Matches 1298; Conservative 0; Mismatches 14; Indels 5; Gaps 5;
Qy 1116 AGCTAGAGGAGTCCAAAGAAACTTGGATTGCTCTTTCTGCACATCTCGGTGGTAGCA 1175
Db 1170 AGCTAGAGGAGTCCAAAGAAACTTGGATTGCTCTTTCTGCACATCTCGGTGGTAGCA 1229
Qy 1176 ATTATTCTCTGAGTGGGAGCTGGACCTGTAAGGATGTTCTTAATTTCTTCTGCTCTGA 1235
Db 1230 ATTATTCTCTGAGTGGGAGCTGGACCTGTAAGGATGTTCTTAATTTCTTCTGCTCTGA 1289
Qy 1236 GAGCAATGCTATGGGCTGCAGATGACTTCTTTAGAGATTTGCTTTTGTAGCAATAGGTA 1295
Db 1290 GAGCAATGCTATGGGCTGCAGATGACTTCTTTAGAGATTTGCTTTTGTAGCAATAGGTA 1349
Qy 1296 GATGGTTTGGTGGGAGCTTGAAGCGGTCAAGTAGTTGGCTACTTTCTGCTTGA 1355
Db 1350 GATGGTTTGGTGGGAGCTTGAAGCGGTCAAGTAGTTGGCTACTTTCTGCTTGA 1409

1356 TCTATTAAATACCTGCGAGCTCTCTCTCTTTTGGGGTGTGTGCGCTGTGATTAGTTCT 1415
 1410 TCTATTAAATACCTGCGAGCTCTCTCTCTTTTGGGGTGTGTGCGCTGTGATTAGTTCT 1469
 1416 GCTTTTAAACCACTCCCTGGATGCAATTTTCCCTCTGCAATTTCCCTCTCTCTCTCTGGA 1475
 1470 GCTTTTAAACCACTCCCTGGATGCAATTTTCCCTCTGCAATTTCCCTCTCTCTCTCTGGA 1529
 1476 GTTCATACAGAGAACTGCACTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1535
 1530 GTTCATACAGAGAACTGCACTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1589
 1536 AATAATCCACCT 1595
 1590 AATAATCCACCT 1649
 1596 TTCTCAGACTATCTG 1655
 1650 TTCTCAGACTATCTG 1709
 1656 ATCAGATGAGACCTTTGATTTATTTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1715
 1710 ATCAGATGAGACCTTTGATTTATTTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1769
 1716 TGACTTGGCTTTTCT 1775
 1770 TGACTTGGCTTTTCT 1829
 1776 CTAGGTAGAGTATGCT 1835
 1830 CTAGGTAGAGTATGCT 1889
 1836 TTTTACAACTCTGCTAGTACATGCT 1895
 1890 TTTTACAACTCTGCTAGTACATGCT 1949
 1896 TCAGTTTAGTATCTCAATCT 1955
 1950 TCAGTTTAGTATCTCAATCT 2009
 1956 CTTCTCCAACT 2015
 2010 CTTCTCCAACT 2069
 2016 TCTAACTCTGCCATCT 2075
 2070 TCTAACTCTGCCATCT 2129
 2076 ATATTAGCGTTGGTGGAGATTTTCT 2135
 2130 ATATTAGCGTTGGTGGAGATTTTCT 2189
 2136 CTTGAGATAATCCACTAAGAAATTTTCT 2195
 2190 CTTGAGATAATCCACTAAGAAATTTTCT 2249
 2196 ATTATCACTGTGCAAGGAAATTTTCT 2255
 2250 ATTATCACTGTGCAAGGAAATTTTCT 2309
 2256 CATCAGAGAGATTTAAATTTTCT 2314
 2310 CATCAGAGAGATTTAAATTTTCT 2369
 2315 A-CAAAAGGAGTGTGTTTATTTAAT-GTCTAGAGTGAAGAAATGCGATA-GAACAATGTA- 2370
 2370 ACCCAAGGAGTGTGTTTATTTAATGCTCTAGGATGAAGAAATGCGATAGCAATGTAAG 2429
 2371 GTACTTCTAATACTAGTAATACATGATTGTTAGTCTAATATCTGTAATATTAATAT 2427
 2430 GTACTTCTAATACTAGTAATACATGATTGTTAGTCTAATATCTGTAATATTAATAT 2486

RESULT 5
 ABX63519
 ID ABX63519 standard; cDNA; 2068 BP.
 XX
 AC ABX63519;
 XX
 DT 26-FEB-2003 (first entry)
 XX
 DE Human cDNA #519 differentially expressed in activated vascular tissue.
 XX
 KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
 KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
 KW gene therapy; vascular disease; cancer; coronary; artery disease;
 KW hypertension; diabetes; pre-eclampsia; restenosis;
 KW ischaemia-reperfusion injury; stroke.
 XX
 OS Homo sapiens.
 XX
 PN US2002137081-A1.
 XX
 PD 26-SEP-2002.
 XX
 PF 08-JAN-2002; 2002US-00044090.
 XX
 PR 28-JUL-2000; 2000US-0222469P.
 PR 08-JAN-2001; 2000US-0260483P.
 XX
 PA (BAND/) BANDMAN O.
 XX
 PI Bandman O;
 XX
 PS WPI; 2003-110597/10.
 XX
 CC This invention relates to a combination comprising several cDNAs that are
 CC differentially expressed in activated vascular tissue. The invention also
 CC discloses a high throughput method for detecting differentially expressed
 CC cDNAs in a sample. The cDNAs of the invention may have
 CC antiatherosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
 CC gynaecological; vasotropic and cerebroprotective activities and may be
 CC used in gene therapy. The cDNAs of the invention may be used in a high-
 CC throughput methods for detecting differential expression of one or more
 CC cDNAs in a sample, or screening several molecules or compounds to
 CC identify a molecule or compound that specifically binds a cDNA of the
 CC invention. A protein encoded by the cDNA may be used to screen several
 CC molecules or compounds to identify a ligand that specifically binds to
 CC the protein, or to produce or purify an antibody to the protein that can
 CC be used to detect a protein in a sample or purify a natural or
 CC recombinant protein from a sample. The nucleotides may be useful for
 CC diagnosing, staging, treating, or monitoring the progression of treatment
 CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
 CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
 CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
 CC genetic or gene expression analysis of several new nucleic acid
 CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
 CC diagnosing pre-pathologic disorders, and chronic or acute diseases
 CC associated with abnormalities in the expression, amount or distribution
 CC of the protein. The present sequence represents a cDNA of the invention
 CC that is differentially expressed in activated vascular tissue. Note: The
 CC sequence data for this patent did not form part of the specification, but
 CC was obtained in electronic format directly from USPTO at
 CC http.segdata.uspto.gov/sequence.html?DocID=20020137081
 XX
 SQ Sequence 2068 BP; 657 A; 338 C; 452 G; 621 T; 0 U; 0 Other;

Query Match 18.4%; Score 456.6; DB 7; Length 2068;
 Best Local Similarity 95.2%; Pred. No. 7.8e-93;

Matches 471; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 801 AATATTTTGAATTAATATCTTGGCTTATGCTTACCTTACAGTTTAAATCCCATGGAT 860

Db 1283 ACATATTTCTATCTTGAACGGTAACATTCGTATAGATACCACTTTATCCCATGGAT 1342

QY 861 CAATCAAAATTAATCATCATGACTACATTTGATTTCCCATCGCTGAAGCAGAGATTCATT 920

Db 1343 CAATCAAAATTAATCATCATGACTACATTTGATTTCCCATCGCTGAAGCAGAGATTCATT 1402

QY 921 GTGTGGCATTTGATTTGATCCGAGCTTATTCATATCTTCTCTCAGATGATAGTAA 980

Db 1403 GTGTGGCATTTGATTTGATCCGAGCTTATTCATATCTTCTCTCAGATGATAGTAA 1462

QY 981 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGGCTGTGGTATCATGTGGCTTTGCTCA 1040

Db 1463 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGGCTGTGGTATCATGTGGCTTTGCTCA 1522

QY 1041 CTCATGTGGATGATGATGATTTGATTTCAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100

Db 1523 CTCATGTGGATGATGATTTGATTTCAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1582

QY 1101 AGCTGTGAGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1160

Db 1583 AGCTGTGAGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1642

QY 1161 TCTCGTGCTTACGAAATTTATTCCTCTGATGGAGCTGGACCTGTAAAGAGATTTCTAA 1220

Db 1643 TCTCGTGCTTACGAAATTTATTCCTCTGATGGAGCTGGACCTGTAAAGAGATTTCTAA 1702

QY 1221 TCTTTCTGCTCTGAGAGCAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTT 1280

Db 1703 TCTTTCTGCTCTGAGAGCAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTT 1762

QY 1281 TTGAGCAAAATAGGTA 1295

Db 1763 TTGAGCAAAATAGGTA 1777

RESULT 6

AA333941

ID AAX33941 standard; DNA; 1335 BP.

AC AAX33941;

DT 30-JUN-1999 (first entry)

XX Human HCMV inducible gene, SEQ ID NO 6.

DE HCMV inducible gene; cig; human; human cytomegalovirus; interferon;

KW anti-viral therapy; anti-HCMV therapy; detection; diagnosis;

KW drug screening; ds.

XX Homo sapiens.

OS WO9113075-A2.

XX 18-MAR-1999.

XX 08-SEP-1998; 98MO-US018638.

XX 08-SEP-1997; 97US-0058180P.

XX 22-SEP-1997; 97US-0059725P.

XX (UYPR-) UNIV PRINCETON.

XX Zhu H, Cong J, Schenk T;

XX WPI; 1999-243729/20.

XX P-PSDB; AAY05371.

XX New isolated human genes to obtain agents for antiviral therapy,

PT particularly anti-HCMV therapy.

XX Claim 2; Page 103-104; 184pp; English.

XX This sequence represents a human gene of the invention, that is induced

CC to express by both HCMV and interferon (IFN), designated HCMV-inducible

CC genes (cig or cigs). The invention also relates to genes that are

CC repressed in the presence of HCMV infection, designated HCMV-repressible

CC genes (crg or crgs). The products can be used to obtain agents which can

CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can

CC also be used for the development of drugs that would allow for higher

CC dosage IFN treatments without the concomitant toxicity normally

CC associated with administering high levels of IFN. The products can also

CC be used for detection, diagnosis and drug screening

XX

SQ Sequence 1335 BP; 409 A; 232 C; 311 G; 383 T; 0 U; 0 Other;

Query Match 18.3%; Score 455; DB 2; Length 1335;

Best Local Similarity 94.9%; Pred. No. 1.6e-92;

Matches 470; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 801 AATATTTTGAATTAATATCTTGGCTTATGCTTACCTTACAGTTTAAATCCCATGGAT 860

Db 797 ACATATTTCTATCTTGAACGGTAACATTCGTATAGATACCACTTTATCCCATGGAT 856

QY 861 CAATCAAAATTAATCATCATGACTACATTTGATTTCCCATCGCTGAAGCAGAGATTCATT 920

Db 857 CAATCAAAATTAATCATCATGACTACATTTGATTTCCCATCGCTGAAGCAGAGATTCATT 916

QY 921 GTGTGGCATTTGATTTGATCCGAGCTTATTCATATCTTCTCTCAGATGATAGTAA 980

Db 917 GTGTGGCATTTGATTTGATCCGAGCTTATTCATATCTTCTCTCAGATGATAGTAA 976

QY 981 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGGCTGTGGTATCATGTGGCTTTGCTCA 1040

Db 977 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGGCTGTGGTATCATGTGGCTTTGCTCA 1036

QY 1041 CTCATGTGGATGATGATTTGATTTCAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100

Db 1037 CTCATGTGGATGATGATTTGATTTCAAAAGGTGACCTTATAGAAATAGAGAGATGTG 1096

QY 1101 AGCTGTGAGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1160

Db 1097 AGCTGTGAGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1156

QY 1161 TCTCGTGCTTACGAAATTTATTCCTCTGATGGAGCTGGACCTGTAAAGAGATTTCTAA 1220

Db 1157 TCTCGTGCTTACGAAATTTATTCCTCTGATGGAGCTGGACCTGTAAAGAGATTTCTAA 1216

QY 1221 TCTTTCTGCTCTGAGAGCAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTT 1280

Db 1217 TCTTTCTGCTCTGAGAGCAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTT 1276

QY 1281 TTGAGCAAAATAGGTA 1295

Db 1277 TTGAGCAAAATAGGTA 1291

RESULT 7

ABN59995

ID ABN59995 standard; cDNA; 1679 BP.

XX AC ABN59995;

XX AC ABN59995;

XX 28-JUN-2002 (first entry)

XX Novel human coding sequence SEQ ID NO: 406.

XX Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;

XX antiinfectility; cerebroprotective; cytostatic; rheumatic; gene therapy;

XX neuroprotective; antiparkinsonian; protein therapy; EST;

XX expressed sequence tag; gene; ss.

OS Homo sapiens.

Db 977 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAACGCTGGTGTGTACATGTGGCTTTTGCTCA 1036
Qy 1041 CTCATGTGGATAGCATGATTGATTACAAAGGTGACCTTATAGAAATAGAGATGTG 1100
Db 1037 CTCATGTGGATAGCATGATTGATTACAAAGGTGACCTTATAGAAATAGAGATGTG 1096
Qy 1101 AGCTGTGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTGGATTGCTCTTCTGACA 1160
Db 1097' TGCCCTGTGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTGGATTGCTCTTCTGACA 1156
Qy 1161 TCTCGGTGTGATGCAATATTCTCTGAGTGGAGCTGGACCCCTGTAAGGATGTTCTAA 1220
Db 1157 TCTCGGTGTGATGCAATATTCTCTGAGTGGAGCTGGACCCCTGTAAGGATGTTCTAA 1216
Qy 1221 TTCCTTCTGCTCTGAGCAAGATGCTATGSGCTGAGATGACTTCTTAGAGGATTTGCCTT 1280
Db 1217 TTCCTTCTGCTCTGAGCAAGATGCTATGSGCTGAGATGACTTCTTAGAGGATTTGCCTT 1276
Qy 1281 TTGAGCAAAATAGGTA 1295
Db 1277 TTGAGCAAAATAGGGA 1291

RESULT 9
AAS60407
ID AAS60407 standard; cDNA; 567 BP.
XX AAS60407;
XX
XX 29-JAN-2002 (first entry)
XX
XX Human cancer agent-sensitve marker #138.
XX
XX Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
XX squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
XX lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
XX Hodgkin's disease; glioma; ss.
XX
XX Homo sapiens.
XX
XX WO200179556-A2.
XX
XX 25-OCT-2001.
XX
XX 13-APR-2001; 2001WO-US012132.
XX
XX 14-APR-2000; 2000US-0197538P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Brown JL, Bolt A, Van Huffel C;
XX WPI; 2001-602933/68.
XX
XX Novel nucleic acid, used as a marker to determine the effectiveness of
XX using TAXOL to treat cancer cell growth in individuals.
XX
XX Claim 1; Page 211; 527pp; English.

XX The invention relates to 1046 novel nucleic acids which are used as
XX markers for determining the sensitivity of a cancer cell to the
XX anticancer agent TAXOL. Cancer cells can be treated with TAXOL when they
XX are shown to express one of the 242 sensitivity markers or the cells are
XX shown not to express one of the 804 resistance markers. The methods can
XX be used to determine the effectiveness of TAXOL in the treatment of
XX cancer cell growth in an individual. The markers can be used as targets
XX in developing anti-cancer agents such as chemotherapeutic compounds. The
XX markers can also be used as targets in developing treatments for cancer,
XX particularly those cancers which display resistance to agents and exhibit
XX expression of the markers. The anticancer agents developed by the novel
XX method can be used to treat cancer. Probes based on the markers can be
XX used to detect transcripts or genomic sequences corresponding to the
XX markers, in the identification of cells or tissues which mis-express the

CC protein. Cancers which may be targeted include carcinoma (e.g. squamous
CC cell carcinoma), sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic
CC leukaemia), lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's
CC disease and tumours (e.g. Glioma). The present sequence is one of the
XX 1046 novel cancer cell markers
SQ Sequence 567 BP; 128 A; 102 C; 116 G; 220 T; 0 U; 1 Other;
Query Match 16.7%; Score 415.4; DB 4; Length 567;
Best Local Similarity 98.4%; Pred. No. 1e-83;
Matches 419; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1116 AGCTAGAGGAAGTCCAAAGAAACTTGGATTGCTCTTCTGACATCTCGGTGGTTAGCA 1175
Db 142 AGCTAGAGGAAGTCCAAAGAAACTTGGATTGCTCTTCTGACATCTCGGTGGTTAGCA 201
Qy 1176 ATTATTCCTCTGAGTGGAGCTGGACCCCTGTAAGGATGTTCTAATTTCTTCTGCTCTGA 1235
Db 202 ATTATTCCTCTGAGTGGAGCTGGACCCCTGTAAGGATGTTCTAATTTCTTCTGCTCTGA 261
Qy 1236 GACGAATGCTATGSGCTGCAGATGACTTCTTAGAGGATTTGCCCTTTGAGCAAAATAGGTA 1295
Db 262 GACGAATGCTATGSGCTGCAGATGACTTCTTAGAGGATTTGCCCTTTGAGCAAAATAGGTA 321
Qy 1296 GATGGTTGGTGTGAGAGCTTGGAAAGCTGAGGTGAGTGGCTACTTTCTGCTTGGGA 1355
Db 322 GATGGTTGGTGTGAGAGCTTGGAAAGCTGAGGTGAGTGGCTACTTTCTGCTTGGGA 381
Qy 1356 TCTATTAAATACCTGGCAGCTCTCTGTTTTTGTGGTGTGTTGCCCTGTGATTAGTTCT 1415
Db 382 TCTATTAAATACCTGGCAGCTCTCTGTTTTTGTGGTGTGTTGCCCTGTGATTAGTTCT 441
Qy 1416 GCTTTTAAACCACTCCCTGGATGCAATTTTCCCTCCTTGCATTTCCCTTTTCTGCTGGA 1475
Db 442 GCTTTTAAACCACTCCCTGGATGCAATTTTCCCTCCTTGCATTTCCCTTTTCTGCTGGA 501
Qy 1476 GTTCATCTAGAGAACTGCACTATGTTTCCCTTTTGTCTTGTGAGATGAAAGTTTTAA 1535
Db 502 AGTCATCTAGAGAACTGCACTATGTTTCCCTTTTGTCTTGTGAGATGAAAGTTTTAA 561
Qy 1536 AATAAT 1541
Db 562 AATAAT 567

RESULT 10
ABL80943
ID ABL80943 standard; cDNA; 482 BP.
XX ABL80943;
XX
XX 17-MAY-2002 (first entry)
XX
XX Human ovarian cancer related cDNA clone SEQ ID NO:3921.
XX
XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200192581-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US017756.
XX
XX 26-MAY-2000; 2000US-0207484P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
XX

PS Claim 1; Page 3193-3194; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention. N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922

XX SQ Sequence 977 BP; 353 A; 158 C; 202 G; 259 T; 0 U; 5 Other;

Query Match 12.7%; Score 315.4; DB 4; Length 977;
Best Local Similarity 82.1%; Pred. No. 4.2e-61;
Matches 372; Conservative 1; Mismatches 76; Indels 4; Gaps 1;

QY 843 AGTTAATCCATGGGAATCAATCAATTAATCAATCATCATGACATACATGATTCCTCCATCGC 902
Db 202 AGATCCACCTAGAAAGGGGAAGCTNTTTCNGGGAACCACTTTATACCCACAC 261
QY 903 TGAAGGACAGAAATCAATTTGTTGGGCAATTTGATTTGATGCGAGCTCTATTCATTTCT 962
Db 262 AAAAATATAGCATGAGCTGTGT-----TTTAGAGGAGATAGGGTGCACCAACCAATTTCACT 317
QY 963 CCTCTCAGATGATAGTAAGATCAAAAGAAATTCGAAGGAGTTGGTAAACGCTGGTGGG 1022
Db 318 CCTCTCAGATGATAGTAAGATCAAAAGAAATTCGAAGGAGTTGGTAAACGCTGGTGGG 377
QY 1023 TACATGTGGCTTTGCTCCTCATCTGATGAGTAGATGATTTGATCAAAAGTGACCTTA 1082
Db 378 TACATGTGGCTTTTCTCCTCATCTGATGAGTAGATGATTTGATCAAAAGTGACCTTA 437
QY 1083 TAGAATATAGAGATGTGAGCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAACTTG 1142
Db 438 TAGAATATAGAGATGTGAGCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAACTTG 497
QY 1143 GATTGCTCTTTCTGACATCTCGTGGTGTAGCAATTTATTCCTCTGAGTGGGAGCTGGACC 1202
Db 498 GATTGCTCTTTCTGACATCTCGTGGTGTAGCAATTTATTCCTCTGAGTGGGAGCTGGACC 557
QY 1203 CTGTAAGGATGTTCTAATTTCTTCTGCTGAGACGAATGCTATGGCTGCAGATGACT 1262
Db 558 CTGTAAGGATGTTCTAATTTCTTCTGCTGAGACGAATGCTATGGCTGCAGATGACT 617
QY 1263 TCTTAGAGGATTTGCCCTTTTGAGCAATAGGTA 1295
Db 618 TCTTAGAGGATTTGCCCTTTTGAGCAATAGGTA 650

RESULT 14
AA527231
ID AA527231 standard; cDNA; 3276 BP.
XX AC AA527231;
XX DT 07-NOV-2001 (first entry)
XX DE cDNA encoding novel signal transduction pathway protein, Seq ID 266.
XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;

KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
XX acquired immune deficiency syndrome.
OS Homo sapiens.
XX WO200154733-A1.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001312.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 06-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.

```
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0255719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX P-PSDB; AAU17314.
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders.
XX Claim 1; SEQ ID NO 266; 880pp; English.
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I) and the antibody to (I) are useful for
XX diagnosing, preventing and treating diseases including immune system
XX disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
XX disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
XX transplant rejections and graft versus host disease, infectious diseases
XX (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
XX other blood-related disorders (sickle cell anaemia), myeloproliferative
XX disorders, primary haematopoietic disorders, hyperproliferative disorders
XX (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
XX Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
XX (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
XX glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
XX respiratory disorders, dermatological disorders, in wound healing,
XX epithelial cell proliferation, endocrine disorders (e.g. Addison's
XX disease), reproductive system disorders, gastrointestinal disorders
XX (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
XX B-cell responsiveness to pathogens, activators of T-cells, to induce
XX higher affinity antibodies, and as a means to induce tumour proliferation
XX in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
XX AAS27850 represent novel signal transduction pathway protein coding
XX sequences and PCR primers of the invention
Query Match 12.5%; Score 311; DB 4; Length 3276;
Best Local Similarity 95.5%; Pred. No. 6e-60;
Matches 320; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 2150 CTAAGGAATATTTTGTGTTCTTTCTCAGGGAATCTAAGGAGGAATATATCACTGTGC 2209
DB 2923 CTTAGAGGATTTGCCCTTTTGAGCAATAGGGAATCTAAGGAGGAATATCACTGTGC 2582
QY 2210 ACAAGGAAAAATAGATATGTGAAGGTTTCAGTAAATTTCTCATCATCACAAGAGNTT 2269
DB 2983 ACAAGGAAAAATAGATATGTGAAGGTTTCAGTAAATTTCTCATCATCACAAGAGNTT 3042
QY 2270 AARATTCAGAAAGGAGAAACACAGACCAAGAGAGTATCTAAGACCAAGAGGATGTGT 2329
DB 3043 AARATTCAGAAAGGAGAAACACAGACCAAGAGAGTATCTAAGACCAAGAGGATGTGT 3102
QY 2330 TTTATTATGCTTAGGATGAAGAAATGTCATAGAACATTGTAGTACTTGTAAATAACTAGA 2389
DB 3103 TTTATTATGCTTAGGATGAAGAAATGTCATAGAACATTGTAGTACTTGTAAATAACTAGA 3162
QY 2390 AATAACATGATTTAGTCATAATTTGTCAAAAATAATAATTTTCTTGGATTATGTTTC 2449
DB 3163 AATAACATGATTTAGTCATAATTTGTCAAAAATAATAATTTTCTTGGATTATGTTTC 3222
QY 2450 TGTATCTGTGAAAAATAAAATTTCTTATAAACTC 2484
DB 3223 TGTATCTGTGAAAAATAAAATTTCTTATAAACTC 3257
```


RESULT 15

AD93409
ID AD93409 standard; cDNA; 3276 BP.
XX AC AD93409;
XX DT
XX DE Human cDNA encoding a novel protein #256.
XX DE
XX ss; gene; human; autoimmune disease; Parkinson's disease; silicosis;
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
KW immunosuppressive agent; adjuvant; enhance immune response;
KW higher affinity antibody induction;
KW increased serum immunoglobulin concentration.
XX OS
XX Homo sapiens.
XX FN US2002168711-A1.
XX PD
XX PD 14-NOV-2002.
XX PF
XX PF 17-JAN-2001; 2001US-00764868.
XX PR
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 11-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 14-JUL-2000; 2000US-0217496P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 22-AUG-2000; 2000US-0225758P.
XX PR 22-AUG-2000; 2000US-0226688P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
XX PR 05-SEP-2000; 2000US-0229509P.
XX PR 05-SEP-2000; 2000US-0229513P.
XX PR 08-SEP-2000; 2000US-0231413P.
XX PR 21-SEP-2000; 2000US-0234223P.
XX PR 21-SEP-2000; 2000US-0234274P.
XX PR 27-SEP-2000; 2000US-0235834P.
XX PR 29-SEP-2000; 2000US-0236327P.
XX PR 29-SEP-2000; 2000US-0236367P.
XX PR 29-SEP-2000; 2000US-0236368P.
XX PR 29-SEP-2000; 2000US-0236369P.
XX PR 02-OCT-2000; 2000US-0236370P.
XX PR 02-OCT-2000; 2000US-0236802P.
XX PR 02-OCT-2000; 2000US-0237037P.
XX PR 02-OCT-2000; 2000US-0237038P.
XX PR 02-OCT-2000; 2000US-0237039P.
XX PR 02-OCT-2000; 2000US-0237040P.
XX PR 13-OCT-2000; 2000US-0239935P.
XX PR 20-OCT-2000; 2000US-0240960P.
XX PR 20-OCT-2000; 2000US-0241785P.
XX PR 20-OCT-2000; 2000US-0241809P.
XX PR 01-NOV-2000; 2000US-0244617P.
XX PR 17-NOV-2000; 2000US-0245299P.
XX PR 08-DEC-2000; 2000US-0251856P.
XX PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-719985/68.
DR P-PSDB; ADB94022.
XX
XX New isolated polypeptide useful for diagnosing and treating
PT immunosuppressive conditions such as autoimmune disease and Parkinson's
PT disease.
XX
XX Claim 3; SEQ ID NO 266; 345pp; English.
XX
XX The invention relates to an isolated polypeptide. The polypeptide is
CC useful for diagnosing a pathological condition or a susceptibility to a
CC pathological condition in a subject, by determining the presence or
CC amount of expression of the polypeptide in a biological sample and
CC diagnosing a pathological condition or a susceptibility to a pathological
CC condition based on the presence or amount of expression of the
CC polypeptide. The polypeptide is also useful for identifying a binding
CC partner to the polypeptide, which involves contacting the polypeptide
CC with a binding partner and determining whether the binding partner
CC effects an activity of the polypeptide. The polypeptide or the nucleic
CC acid encoding the polypeptide is useful for preventing, treating, or
CC ameliorating a medical condition, which involves administering the
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
CC is useful for diagnosing a pathological condition or a susceptibility to
CC a pathological condition in a subject, which involves determining the
CC presence or absence of a mutation in the nucleic acid, and diagnosing a
CC pathological condition or susceptibility to a pathological condition
CC based on the presence or absence of the mutation. The polypeptide, the
CC nucleic acid and an antibody to the polypeptide are useful for treating
CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
CC the nucleic acid and the antibody are useful as immunosuppressive agents,
CC as adjuvants to enhance immune responses, and as agents to induce higher
CC affinity antibodies and increase serum immunoglobulin concentrations. The
CC present sequence represents cDNA encoding a novel human protein. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format direct from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=2002016871.
XX
XX Sequence 3276 BP; 1075 A; 495 C; 669 G; 1037 T; 0 U; 0 Other;
Query Match 12.5%; Score 311; DB 9; Length 3276;
Best Local Similarity 95.5%; Pred. No. 6e-60; Indels 0; Gaps 0;
Matches 320; Conservative 0; Mismatches 15;
QY 2150 CTAAGAATATTTTGTGTTCTTTCTCAGGAATCTAAGGAGGAAATATCAACTGTGC 2209
DB 2923 CTTAGAGGATTTGCTTTTGAGCAATAGGAATCTAAGGAGGAAATATCAACTGTGC 2982
QY 2210 ACAAGGAAAAATAGATGTTGAAGGTTCAAGTAAATTTCTTCACATCACAGAAGATT 2269
DB 2983 ACAAGGAAAAATAGATGTTGAAGGTTCAAGTAAATTTCTTCACATCACAGAAGATT 3042
QY 2270 AAAATTCAGAAAGGAGAAAAACACAGACCAAGAGAGATATCTAAGACCAAAAGGATGTGT 2329
DB 3043 AAAATTCAGAAAGGAGAAAAACACAGACCAAGAGAGATATCTAAGACCAAAAGGATGTGT 3102
QY 2330 TTTTATTAATGCTAGGATCAGAAATGTCATAGACATTTAGTACTTCTTAATTAAGTACA 2389
DB 3103 TTTTATTAATGCTAGGATCAGAAATGTCATAGACATTTAGTACTTCTTAATTAAGTACA 3162
QY 2390 AATAACATGATTTAGTCATTAATTTGTGAAAAATAATAAATTTCTTGGATTTATGTTTC 2449
DB 3163 AATAACATGATTTAGTCATTAATTTGTGAAAAATAATAAATTTCTTGGATTTATGTTTC 3222
QY 2450 TGTATCTGTGAAAAATAATAAATTTCTTATAAACTC 2484

Db 3223 TGTATCTGTGAAAAATAAATTTCTTATAAACTC 3257

Search completed: April 25, 2004, 08:46:17
Job time : 665.01 secs

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1369	55.1	3062	9	HSM805959	BS537508 Homo sapi
2	1369	55.1	151265	9	AC104837	AC104837 Homo sapi
3	1235.2	49.7	2527	6	AX285019	AX285019 Sequence
4	456.6	18.4	1714	9	BC022870	BC022870 Homo sapi
5	455	18.3	1679	6	AX405991	AX405991 Sequence
6	448.6	18.1	1062	6	E01792	E01792 DNA sequence
7	448.6	18.1	1062	6	E01797	E01797 DNA encodin
8	448.6	18.1	1332	6	E01791	E01791 DNA sequenc
9	448.6	18.1	1332	6	E01975	E01975 DNA encodin
10	448.6	18.1	1665	9	CHPP44	D90034 Pan troglod
11	448.6	18.1	1667	6	E01978	E01978 DNA sequenc
12	448.6	18.1	1667	6	E01978	AX284603 Sequence
13	415.4	16.7	567	6	AX284603	D28915 Homo sapien
14	364	14.7	483	9	D2890858	AX284296 Sequence
15	331.2	13.3	580	6	AX284296	AX299405 Mus muscu
16	283.2	11.4	2797	10	MMU239405	AX378997 Sequence
17	272	11.0	272	6	AX378997	D28914 Homo sapien
18	252	10.1	372	9	D2890857	AR240306 Sequence
19	240	9.7	240	6	AR240306	D28912 Homo sapien
20	232.4	9.4	341	9	D2890855	BC026901 Mus muscu
21	223.8	9.0	1793	10	BC026901	AB000115 Homo sapi
22	196.8	7.9	2058	9	AB000115	BD21210 Human cyt
23	196.8	7.9	2058	6	BD21210	BC015932 Homo sapi
24	195.2	7.9	2058	9	BC015932	AL832618 Homo sapi
25	195.2	7.9	4757	9	HSM803928	BC024930 Mus muscu
26	165.6	6.7	2228	10	BC024930	AC123055 Mus muscu
27	160	6.4	187323	10	AC123055	AF336221 Mus muscu
28	158	6.4	2279	10	AF336221	AC099264 Rattus no
29	143.6	5.8	252914	2	AC099264	D28913 Homo sapien
30	106.8	4.3	288	9	D2890856	AX35818 Sequence 14
31	101.2	4.1	489	6	AX35818	166494 Sequence 14
32	90.6	3.6	7218	6	166494	AC108332 Rattus no
33	84	3.4	218015	2	AC108332	AP003165 Homo sapi
34	72	2.9	112843	9	AP003165	AC093218 Homo sapi
35	71.6	2.9	105360	2	AC093218	AC021318 Homo sapi
36	71.6	2.9	198784	2	AC021318	AF538053 Monosiga
37	69.2	2.8	76588	3	MBREV	AC005376 Homo sapi
38	66	2.7	82291	9	AC005376	AL592425 Homo sapi
39	65.2	2.6	169318	2	AL592425	AC083979 Homo sapi
40	65.2	2.6	173660	9	AC083979	AL583829 Human DNA
41	64.6	2.6	49573	9	AL583829	AC138073 Homo sapi
42	64.6	2.6	93791	2	AC138073	AC027069 Homo sapi
43	64.2	2.6	145063	2	AC027069	AL356292 Human DNA
44	64.2	2.6	154088	9	AL356292	AC074085 Homo sapi
45	63.4	2.6	155280	9	AC074085	

ALIGNMENTS

RESULT 1	HSM805959	HSM805959	3062 bp	mRNA	linear	PRI 17-JUN-2003
LOCUS	Homo sapiens	Homo sapiens mRNA; cDNA DKFZp686L21223 (from clone DKFZp686L21223).				
DEFINITION	BS537508					
ACCESSION	BS537508					
VERSION	BS537508.1	GI:31873567				
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 3062)					
AUTHORS	Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,					
	Robo G., Han M. and Wiemann S.					
TITLE	Direct Submission					

Pred. No. is the number of results predicted by chance to have a

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 6540.9 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-18

Perfect score: 2484

Sequence: 1 gttattctgtattataac.....ataaattctataaaactc 2484

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.pl.*
- 8: gb.pr.*
- 9: gb.ro.*
- 10: gb.sts.*
- 11: gb.sy.*
- 12: gb.un.*
- 13: gb.vi.*
- 14: em.ba.*
- 15: em.fun.*
- 16: em.in.*
- 17: em.mu.*
- 18: em.or.*
- 19: em.ov.*
- 20: em.pat.*
- 21: em.ph.*
- 22: em.pl.*
- 23: em.ro.*
- 24: em.sts.*
- 25: em.un.*
- 26: em.vi.*
- 27: em.htg.hum.*
- 28: em.htg.inv.*
- 29: em.htg.other.*
- 30: em.htg.mus.*
- 31: em.htg.pin.*
- 32: em.htg.mam.*
- 33: em.htg.rod.*
- 34: em.htg.vrt.*
- 35: em.sv.*
- 36: em.higo.hum.*
- 37: em.higo.mus.*
- 38: em.higo.other.*
- 39: em.higo.mus.*
- 40: em.higo.mus.*
- 41: em.higo.mus.*

1776	Qy	CTAGGTAGAGGTATGCTCCTTTTATGTAAATCGCCACCGCATTTAGTATTACATAAACTTTC	1835
2325	Db	CTAGGTAGAGGTATGCTCCTTTTATGTAAATCGCCACCGCATTTAGTATTACATAAACTTTC	2384
1836	Qy	TTTTAACAACTGTGTCATAGTACATGCTGCTGTGTTCCATTTAGAGATTTGACAGAGTT	1895
2385	Db	TTTTAACAACTGTGTCATAGTACATGCTGCTGTGTTCCATTTAGAGATTTGACAGAGTT	2444
1896	Qy	TCAGTTTATGATATCTCAAACTCTTATTTTATGTGCTTGGGAAATCAATTCAGAATATCACAT	1955
2445	Db	TCAGTTTATGATATCTCAAACTCTTATTTTATGTGCTTGGGAAATCAATTCAGAATATCACAT	2504
1956	Qy	CCTCTCAATTTCTCTTTACTCAAAATTTGCTGGGAAATCTCATGTTACTTAACCTTTGTTC	2015
2505	Db	CCTCTCAATTTCTCTTTACTCAAAATTTGCTGGGAAATCTCATGTTACTTAACCTTTGTTC	2564
2016	Qy	TCTAACTCTGCCATCTTGGTTTCCCAATCCCTTCTCTTCTCATGTGACGTGTCTCCTA	2075
2565	Db	TCTAACTCTGCCATCTTGGTTTCCCAATCCCTTCTCTTCTCATGTGACGTGTCTCCTA	2624
2076	Qy	ATATTAGGTTGGTTGAGATTTTTCAGTGTGCCAATATTTCTCTCTCCCTCTGTTGCCCTTT	2135
2625	Db	ATATTAGGTTGGTTGAGATTTTTCAGTGTGCCAATATTTCTCTCTCCCTCTGTTGCCCTTT	2684
2136	Qy	CCTGAGATAATCCACTAAGAAATATTTTGTGTTCTTTTCTCAGGGAATCTAAGGGAGGAA	2195
2685	Db	CCTGAGATAATCCACTAAGAAATATTTTGTGTTCTTTTCTCAGGGAATCTAAGGGAGGAA	2744
2196	Qy	ATTATCAACTGTGCACAGGAGGAAAAATAGATATGTGAAAGGTTTCAGCTAAATTTCTCTCA	2255
2745	Db	ATTATCAACTGTGCACAGGAGGAAAAATAGATATGTGAAAGGTTTCAGCTAAATTTCTCTCA	2804
2256	Qy	CATCAGAGAATTTAAATTCGAAAGGAGGAAAAACACAGACCAGAGAGTATCTTAAGA	2315
2805	Db	CATCAGAGAATTTAAATTCGAAAGGAGGAAAAACACAGACCAGAGAGTATCTTAAGA	2864
2316	Qy	CCAAAGGGATGTGTTTATTTAATCTCTAGGATGAAGAAATGTCATAGAACCTTTGTAGTACT	2375
2865	Db	CCAAAGGGATGTGTTTATTTAATCTCTAGGATGAAGAAATGTCATAGAACCTTTGTAGTACT	2924
2376	Qy	TGTAATAACTAGAAATAACATGATTTTGTGCATATTTGTGAAAAATATATATAATTTTTTC	2435
2925	Db	TGTAATAACTAGAAATAACATGATTTTGTGCATATTTGTGAAAAATATATATAATTTTTTC	2984
2436	Qy	TTGGATTATTTGTTCTGTATCTGTGAAAAATAAATTTCTTTATAAAACTC	2484
2985	Db	TTGGATTATTTGTTCTGTATCTGTGAAAAATAAATTTCTTTATAAAACTC	3033

TITLE
JOURNAL
COMMENT

Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
Direct Submission
Submitted (19-JAN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jan 19, 2002 this sequence version replaced gi:17976475.
Genome Center

Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: SC

----- Project Information
Center project name: chr-1
Center clone name: RP4-641G12 (sc0195)

----- Summary Statistics
Sequencing vector: plasmid; 108752; 31% of reads
Sequencing vector: plasmid; 69% of reads
Chemistry: Dye-terminator ET; 75% of reads
Chemistry: Dye-terminator Big Dye; 25% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151265 bases at least Q40
Consensus quality: 151265 bases at least Q30
Consensus quality: 151265 bases at least Q20
Insert size: 151265; sum-of-contigs
Quality coverage: 13.7x in Q20 bases; sum-of-contigs

----- Overlapping Sequences:
5': RP11-369G23 (UWGC:sc0385) AL354946
3': RP11-143H12 (UWGC:sc0641) AC099674

----- Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:
This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
10449	10293	4702	4660	7487	7676
2184	2175	449	<800	5671	5775
8065	8212	512	<800	4181	4233
486	<800	2814	2820	7761	7676

106	<800	1247	1297	5168	5105
4682	4705	4695	4660	1735	1698
606	<800	3373	3381	529	<800
4198	4225	5592	5541	643	<800
3725	3694	2918	2948	1685	1698
5850	6047	4933	4843	981	921
418	<800	12634	12485	1768	1698
575	<800	1681	1612	3918	3844
337	<800	2417	2408	1954	1985
290	<800	765	781	144	<800
1493	1463	5897	5894	1072	1072
1375	1373	2196	2178	83	<800
777	786	1337	1297	2311	2150
2880	2953	671	<800	5786	5775
998	985	784	781	125	<800
119	<800	499	<800	5168	5105
1401	1373	6819	6736	2370	2369
305	<800	2344	2284	6218	6155
1953	1961	15	<800	1401	1387
8699	8762	73	<800	3103	3162
918	985	9	<800	1596	1534
6245	6047	15	<800	2183	2150
1288	1264	4408	4390	1008	1072
4012	3971	507	<800	2306	2150
2789	2842	6659	6736	6663	6646
109	<800	592	<800	2485	2511
791	786	4112	4049	1547	1534
1112	1093	2133	2178	3149	3162
1445	1373	4311	4257	4483	4486
3016	2953	1158	1195	448	<800
7827	7803	7399	7536	3382	3416
1713	1686	211	<800	1229	1237
881	890	3006	2948	6669	6646
5148	5107	6064	6115	6227	6155
2540	2582	1785	1863	2346	2369
407	<800	5272	5233	2369	2369
82	<800	6135	6115	3876	3844

Query Match									
Best Local Similarity 55.1%; Score 1369; DB 9; Length 151265;									
Matches 1369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1116	AGTAGAGAGTCCAAAGAAACCTGGATTGCTCTTTCTGACATCTCGGTGGTTAGCA	1175						
DB	59035	AGTAGAGAGTCCAAAGAAACCTGGATTGCTCTTTCTGACATCTCGGTGGTTAGCA	59094						
QY	1176	ATTATTCTCTGAGTGGAGCTGGACCTGTAAGGATGCTCTAACTTTCTGCTCTGA	1235						
DB	59095	ATTATTCTCTGAGTGGAGCTGGACCTGTAAGGATGCTCTAACTTTCTGCTCTGA	59154						
QY	1236	GACGAATGCTATGGGTGACAGTACTCTTAGAGGATTTGGCTTTTGAGCAATAGGTA	1295						
DB	59155	GACGAATGCTATGGGTGACAGTACTCTTAGAGGATTTGGCTTTTGAGCAATAGGTA	59214						
QY	1296	GATGGTTTGGTGTGGAGCTTGAAGCGGTGAGTGTGGTCTTTCTGCTTGA	1355						
DB	59215	GATGGTTTGGTGTGGAGCTTGAAGCGGTGAGTGTGGTCTTTCTGCTTGA	59274						
QY	1356	TCTATTAAATACCTGCAGCTCTCTCTTTTGTGGTGTGGTCTGATTAAGTTCT	1415						
DB	59275	TCTATTAAATACCTGCAGCTCTCTCTTTTGTGGTGTGGTCTGATTAAGTTCT	59334						
QY	1416	GCCTTTTAAACCACTCCCTGGATGCAATTTCCCTCTTGCAATTTCCCTTTCTCTGA	1475						
DB	59335	GCCTTTTAAACCACTCCCTGGATGCAATTTCCCTCTTGCAATTTCCCTTTCTCTGA	59394						
QY	1476	GTTTCATACAGAACTGCACATATGTTTTCCTTTTGTCTGTGAGATGAAGTTTAA	1535						
DB	59395	GTTTCATACAGAACTGCACATATGTTTTCCTTTTGTCTGTGAGATGAAGTTTAA	59454						
QY	1536	AATAATCCACTCTGCATTTCCACTCTGACATCCCAAGCTGATCCCTGGCTCTT	1595						
DB	59455	AATAATCCACTCTGCATTTCCACTCTGACATCCCAAGCTGATCCCTGGCTCTT	59514						
QY	1596	TTCTCAGACTATGTTCTTTTACTTTGGACCTAGAACTGGATTGCAATGGCTGCTCG	1655						
DB	59515	TTCTCAGACTATGTTCTTTTACTTTGGACCTAGAACTGGATTGCAATGGCTGCTCG	59574						

QY	1656	ATCAGATGAGACCTTTGATTAATTTGCCCTTCTCTTAGGACCTTACACTCCTGCTTTCTT	1715						
DB	59575	ATCAGATGAGACCTTTGATTAATTTGCCCTTCTCTTAGGACCTTACACTCCTGCTTTCTT	59634						
QY	1716	TGACTTTGCCCTTTTGTCTTTCTTCTTCACTTAGTCCCTCTTCACTGATGCTGCTCATG	1775						
DB	59635	TGACTTTGCCCTTTTGTCTTTCTTCTTCACTTAGTCCCTCTTCACTGATGCTGCTCATG	59694						
QY	1776	CTAGGTAGAGGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTACATAAACTTTC	1835						
DB	59695	CTAGGTAGAGGTATGCTCTTTTATGTAATGGCCACCGCATTTAGTATTACATAAACTTTC	59754						
QY	1836	TTTTAACAACTCTGTCATAGTACATGCTCTGTTTCCATTTAGAGATTGACAGAGGTT	1895						
DB	59755	TTTTAACAACTCTGTCATAGTACATGCTCTGTTTCCATTTAGAGATTGACAGAGGTT	59814						
QY	1896	TCAGTTTAGTATCTCAAACTTTTATGTTAGTCTGGGAAATCAATTCAGATATCACAT	1955						
DB	59815	TCAGTTTAGTATCTCAAACTTTTATGTTAGTCTGGGAAATCAATTCAGATATCACAT	59874						
QY	1956	CCTCTCCAACTCTCTTACTCAAAATTTGCTGGGAACTCTCATGTTACTAACTTTGCTTC	2015						
DB	59875	CCTCTCCAACTCTCTTACTCAAAATTTGCTGGGAACTCTCATGTTACTAACTTTGCTTC	59934						
QY	2016	TCTAACTCTCCCATCTGTTTCCCATCCCTTCTTCTCTCATGTTAGTGTGCTCTTA	2075						
DB	59935	TCTAACTCTCCCATCTGTTTCCCATCCCTTCTTCTCTCATGTTAGTGTGCTCTTA	59994						
QY	2076	ATATTAGCGTGTGATGATTTTTCAGTGGTCCCAATATTTCTCTCTGTTGCTTT	2135						
DB	59995	ATATTAGCGTGTGATGATTTTTCAGTGGTCCCAATATTTCTCTCTGTTGCTTT	60054						
QY	2136	CCTGAGATAATCCACTAAGAATATTTGTGTTTCTTCTCAGGGAATCTAAGGAGGAA	2195						
DB	60055	CCTGAGATAATCCACTAAGAATATTTGTGTTTCTTCTCAGGGAATCTAAGGAGGAA	60114						
QY	2196	ATTATCAACTCTGCACAGGAAATATAGATATGTAAGGTTACGTAAATTTCTCTCA	2255						
DB	60115	ATTATCAACTCTGCACAGGAAATATAGATATGTAAGGTTACGTAAATTTCTCTCA	60174						
QY	2256	CATCACAGAGATTAATAATTCAGAAAGGAAACACAGACCAAGAGATCTAAGA	2315						
DB	60175	CATCACAGAGATTAATAATTCAGAAAGGAAACACAGACCAAGAGATCTAAGA	60234						
QY	2316	CCAAAGGATGCTGTTTATTAATGCTTAGGATGAAGAAATGATAGATTTCTAGTACT	2375						
DB	60235	CCAAAGGATGCTGTTTATTAATGCTTAGGATGAAGAAATGATAGATTTCTAGTACT	60294						
QY	2376	TGTAATAAATAGATAAATCAATGATTTAGTCAATAATTTGTGAAATAATAATAATTTTC	2435						
DB	60295	TGTAATAAATAGATAAATCAATGATTTAGTCAATAATTTGTGAAATAATAATAATTTTC	60354						
QY	2436	TTGGATTATGTTCTGATCTGTGAAATAATAATTTCTTATAAACTC	2494						
DB	60355	TTGGATTATGTTCTGATCTGTGAAATAATAATTTCTTATAAACTC	60403						

RESULT 3
AX285019
LOCUS AX285019
DEFINITION Sequence 824 from Patent WO0179556.
ACCESSION AX285019
VERSION AX285019.1 GI:17045707
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lillie, J., Brown, J.L., Bolt, A. and van Hufel, C.
TITLE Novel genes, compositions and methods for the identification, assessment, prevention, and therapy of human cancers

COMMENT

Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 32 Row: n Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361309.

FEATURES

source
1..1714
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:24007 IMAGE:4071404"
/issue_type="Lung"
/clone_lib="NIH MGC 77"
/lab_host="DH105"
/notes="Vector: pDNR-LIB"
1..1714
/gene="IFI44"
/note="synonyms: MTAP44, p44"
/db_xref="LocusID:10561"
90..1424
/codon_start=1
/product="interferon-induced, hepatitis C-associated microtubular aggregate"
/protein_id="AAH22870.1"
/db_xref="GI:11860554"
/db_xref="LocusID:10561"
/translation="MAVTRLTWHLKILQNHFGKRLSLLYKGVHGFNGVLLDRC
CNQPTLTIVSDHIIIGAYAESYQEGKVAIILFALQDTKISWKLGLCPETLFC
CDVTKNSPTNFQIDGRNRKVMIDLKTMENLAQNTISIOYEVFRCDSLDERKI
KGVIELRSLLSALRYEPYVGSVQVQIRILLGLPIGAKSFFNSVFGQHVTHQA
LVGNTTIGSEKRTYTSIRGKQKGLPFLICDSLGLSEKGLCRDDIYILNGNIR
DRYQFNPMESIKNHHYIDPSLKRIHCVAVFVDASSIQYFSSQMIKIKRIQREL
VNAVGVHALLTHVDSMDLTGDLIEIERCEPVRSLKEEVQRKLGFSALSDISVSNT
SSEWELDPVKDVLILSALRMLWAADDLEDLPFEQIGNLREEIINCAQGRK"
141..302
/note="TLDc; Region: domain in TBC and LysM domain containing proteins"
/db_xref="CDD:smart00584"

CDS

gene
1..1679
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
67..1401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35098.1"
/db_xref="GI:21439422"
/translation="MAVTRLTWHLKILQNHFGKRLSLLYKGVHGFNGVLLDRC
CNQPTLTIVSDHIIIGAYAESYQEGKVAIILFALQDTKISWKLGLCPETLFC
CDVTKNSPTNFQIDGRNRKVMIDLKTMENLAQNTISIOYEVFRCDSLDERKI
KGVIELRSLLSALRYEPYVGSVQVQIRILLGLPIGAKSFFNSVFGQHVTHQA
LVGNTTIGSEKRTYTSIRGKQKGLPFLICDSLGLSEKGLCRDDIYILNGNIR
DRYQFNPMESIKNHHYIDPSLKRIHCVAVFVDASSIQYFSSQMIKIKRIQREL
VNAVGVHALLTHVDSMDLTGDLIEIERCEPVRSLKEEVQRKLGFSALSDISVSNT
SSEWELDPVKDVLILSALRMLWAADDLEDLPFEQIGNLREEIINCAQGRK"

misc_feature

141..302
/note="TLDc; Region: domain in TBC and LysM domain containing proteins"
/db_xref="CDD:smart00584"

ORIGIN

Query Match 18.4%; Score 456.6; DB 9; Length 1714;
Best Local Similarity 95.2%; Pred. No. 5.4e-80;
Matches 471; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 801 AAATATTTTGAATTAATATCTTGCTTATGCTACATGTTAATCCCATGAAT 860
DB 886 ACATATTTCTATATCTTGAACGGTAACATTCGTGATAGATACCATGTTAATCCCATGAAT 945
QY 861 CAATCAAAATTAATCATCATGACTACATGATTCCTCCCATCGCTGAAGGACAGAAATTCATT 920
DB 946 CAATCAAAATTAATCATCATGACTACATGATTCCTCCCATCGCTGAAGGACAGAAATTCATT 1005
QY 921 GTGTGGCATTTGATTTGATCCAGCTCTATTCAATCTCTCTCTCAGATGATGATA 980
DB 1006 GTGTGGCATTTGATTTGATCCAGCTCTATTCAATCTCTCTCTCAGATGATGATA 1065
QY 981 AGATCAAAAGAAATTCGAAGGAGATTTGGTAAACGCTGGTGGTACATGTGGCTTGTCTCA 1040
DB 1066 AGATCAAAAGAAATTCGAAGGAGATTTGGTAAACGCTGGTGGTACATGTGGCTTGTCTCA 1125
QY 1041 CTCATGTGGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGATGTG 1100

DB 1126 CTCATGTGGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGATGTG 1185
QY 1101 AGCTGTGAGGTCCAAAGCTAGAGAAAGTCCAAAGAAATCTTGGATTTGCTTTCTTGACA 1160
DB 1186 AGCTGTGAGGTCCAAAGCTAGAGAAAGTCCAAAGAAATCTTGGATTTGCTTTCTTGACA 1245
QY 1161 TCTCGTGTGATGACAAATATTCTCTGAGTGGAGTGGACCTGTAAAGGATGTCTTA 1220
DB 1246 TCTCGTGTGATGACAAATATTCTCTGAGTGGAGTGGACCTGTAAAGGATGTCTTA 1305
QY 1221 TTCTTTCTGCTCTGAGACGAATCTATGGCTCGAGATGACTTCTTAGAGGATTTGCCCT 1280
DB 1306 TTCTTTCTGCTCTGAGACGAATCTATGGCTCGAGATGACTTCTTAGAGGATTTGCCCT 1365
QY 1281 TTGAGCAAAATAGTA 1295
DB 1366 TTGAGCAAAATAGGA 1380

RESULT 5
AX405991 LOCUS 1679 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 406 from Patent WO0222660.
ACCESSION AX405991
VERSION AX405991.1 GI:21439421
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, P., Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 406 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
source
1..1679
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
67..1401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35098.1"
/db_xref="GI:21439422"
/translation="MAVTRLTWHLKILQNHFGKRLSLLYKGVHGFNGVLLDRC
CNQPTLTIVSDHIIIGAYAESYQEGKVAIILFALQDTKISWKLGLCPETLFC
CDVTKNSPTNFQIDGRNRKVMIDLKTMENLAQNTISIOYEVFRCDSLDERKI
KGVIELRSLLSALRYEPYVGSVQVQIRILLGLPIGAKSFFNSVFGQHVTHQA
LVGNTTIGSEKRTYTSIRGKQKGLPFLICDSLGLSEKGLCRDDIYILNGNIR
DRYQFNPMESIKNHHYIDPSLKRIHCVAVFVDASSIQYFSSQMIKIKRIQREL
VNAVGVHALLTHVDSMDLTGDLIEIERCEPVRSLKEEVQRKLGFSALSDISVSNT
SSEWELDPVKDVLILSALRMLWAADDLEDLPFEQIGNLREEIINCAQGRK"

ORIGIN
Query Match 18.3%; Score 455; DB 6; Length 1679;
Best Local Similarity 94.9%; Pred. No. 1.1e-79;
Matches 470; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 801 AAATATTTTGAATTAATATCTTGCTTATGCTACATGTTAATCCCATGAAT 860
DB 863 ACATATTTCTATATCTTGAACGGTAACATTCGTGATAGATACCATGTTAATCCCATGAAT 922
QY 861 CAATCAAAATTAATCATCATGACTACATGATTCCTCCCATCGCTGAAGGACAGAAATTCATT 920
DB 923 CAATCAAAATTAATCATCATGACTACATGATTCCTCCCATCGCTGAAGGACAGAAATTCATT 982
QY 921 GTGTGGCATTTGATTTGATCCAGCTCTATTCAATCTCTCTCTCAGATGATGATA 980
DB 983 GTGTGGCATTTGATTTGATCCAGCTCTATTCAATCTCTCTCTCAGATGATGATA 1042

QY 981 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGGCTGGTGGTACATGTGGCTTTGGCTCA 1040
Db 1043 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGGCTGGTGGTACATGTGGCTTTGGCTCA 1102
QY 1041 CTCATGTGGTAGCATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
Db 1103 CTCATGTGGTAGCATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1162
QY 1101 AGCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1160
Db 1163 AGCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1222
QY 1161 TCTCGGTGGTAGCAATTTATTCCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 1220
Db 1223 TCTCGGTGGTAGCAATTTATTCCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 1282
QY 1221 TTTCTTTCTGCTCTGAGAGCAATGCTATGGCTGAGATGACTTCTTAGAGGATTTGCTT 1280
Db 1283 TTTCTTTCTGCTCTGAGAGCAATGCTATGGCTGAGATGACTTCTTAGAGGATTTGCTT 1342
QY 1281 TTGAGCAAAATAGGTA 1295
Db 1343 TTGAGCAAAATAGGGA 1357

RESULT 6
E01792
LOCUS
DEFINITION DNA sequence coding for non-A non-B hepatitis specific antigen
ACCESSION E01792
VERSION E01792.1 GI:2170045
KEYWORDS JP 1989002576-A/2
SOURCE Pan paniscus (pygmy chimpanzee)
ORGANISM Pan paniscus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1062)
AUTHORS Takahashi,K., Shibui,T., Uchida,M., Kurebayashi,R., Teranishi,Y.,
Nakanishi,S. and Kitamura,N.
TITLE DNA FRAGMENT
JOURNAL Patent: JP 1989002576-A 2 06-JAN-1989;
COMMENT OS Pan paniscus (Chimpanzee)
PN JP 1989002576-A/2
PD 06-JAN-1989
PF 04-JUN-1987 JP 1987140586
PR 31-MAR-1987 JP 87P 78313
PI TAKAHASHI KAZUNOBU, SHIBUI TATSURO, UCHIDA MICHIRU, PI
KUREBAYASHI RIE,
TERANISHI YUTAKA, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
C12N15/00, (C12N15/00, C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue=Liver;
FH Key
FT CDS 1..1062
FT /product='Non-A non-B hepatitis specific antigen protein'
FT Location/Qualifiers
FT 1..1062
FT /organism='Pan paniscus'
FT /mol_type='genomic RNA'
FT /db_xref='taxon:9597'

FEATURES
source
1..1062
Location/Qualifiers
1..1062
/product='Non-A non-B hepatitis specific antigen protein'
/organism='Pan paniscus'
/mol_type='genomic RNA'
/db_xref='taxon:9597'
ORIGIN
Query Match 18.1%; Score 448.6; DB 6; Length 1062;
Best Local Similarity 94.1%; Pred. No. 2.2e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 801 AAATATTTTTGAATTAATATCTTCTTTATCTCTACCTTATATCCATGTTTAAATCCCATGGAAT 860
Db 524 ACATATCCCTACATCTTTGAACGGTAAACATTCGTGATAGATACCAATTTTAAATCCCATGGAAT 583
QY 861 CAATCAAAATTAATCATCATGACTACATGATTTCCCATCGCTGAGGACAGATTCATT 920
Db 584 CAATCAAAATTAATCATCATGACTACATGATTTCCCATCGCTGAGGACAGATTCATT 643
QY 921 GTGTGGCAATTTGATTGATGCCAGCTCTAATCAATACCTTCTCTCTCAGATGATAGTAA 980
Db 644 GTGTGGCAATTTGATTGATGCCAGCTCTAATCAATACCTTCTCTCTCAGATGATAGTAA 703
QY 981 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGGCTGGTGGTACATGTGGCTTTGGCTCA 1040
Db 704 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGGCTGGTGGTACATGTGGCTTTGGCTCA 763
QY 1041 CTCATGTGATGATGATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
Db 764 CTCATGTGATGATGATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGAGATGTG 823
QY 1101 AGCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 1160
Db 824 TGCCCTGTGAGGTCCAAAGCTAGAGGAAGTCCAAAGAAACTTTGGATTTCCTTTCTGACA 883
QY 1161 TCTCGGTGGTAGCAATTTATTCCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 1220
Db 884 TCTCGGTGGTAGCAATTTATTCCTCTGAGTGGAGCTGGACCTGTAAAGGATGTTCTAA 943
QY 1221 TTTCTTTCTGCTCTGAGAGCAATGCTATGGCTGAGATGACTTCTTAGAGGATTTGCTT 1280
Db 944 TTTCTTTCTGCTCTGAGAGCAATGCTATGGCTGAGATGACTTCTTAGAGGATTTGCTT 1003
QY 1281 TTGAGCAAAATAGGTA 1295
Db 1004 TTGAGCAAAATAGGGA 1018
E01977
LOCUS
DEFINITION DNA encoding C-terminal peptide of type nonA nonB hepatitis specific antigen.
ACCESSION E01977.1 GI:2170225
VERSION E01977.1
KEYWORDS JP 1989124387-A/3.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 1062)
AUTHORS Shibui,T., Kurebayashi,R., Uesono,M., Teranishi,Y., Takahashi,K.,
Nakanishi,S. and Kitamura,N.
TITLE MANIFESTATION VECTOR HAVING DNA CODING NON-A NON-B HEPATITIS
SPECIFIC ANTIGEN, TRANSFORMANT AND PRODUCTION OF SAID ANTIGEN
JOURNAL Patent: JP 1989124387-A 3 17-MAY-1989;
COMMENT MITSUBISHI KASEI CORP
OS Pan troglodytes
PN JP 1989124387-A/3
PD 17-MAY-1989
PF 10-NOV-1987 JP 1987283990
PI SHIBUI TATSURO, KUREBAYASHI RIE, UESONO MICHIRU, PI
TERANISHI YUTAKA,
TAKAHASHI KAZUNOBU, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
C12N15/00, A61K39/29, C12N1/20, C12P21/02, (C12P21/02, C12R1:19); CC
strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue type=Liver;
CC *source: clone=PCDCL-1;
CC Feature is identified by experimental;
FH Key Location/Qualifiers
FH

FT misc_feature 1..1049
FT /note="DNA encoding C-terminal peptide of type
FT non-B
FT hepatitis specific antigen".
FEATURES
source
Location/Qualifiers
1..1062
/organism="Pan troglodytes"
/mol_type="genomic RNA"
/db_xref="taxon:9598"
ORIGIN
Query Match 18.1%; Score 448.6; DB 6; Length 1062;
Best Local Similarity 94.1%; Pred. No. 2.2e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 801 AATATATTTTGAATTAATATCTTGTCTTATGCTACCTTACAGTTAATCCATGGAA 860
Db 524 ACATATCCTACATCTTGAACGGTAACATTCGTGTAGATACCAAGTTAATCCATGGAA 583
QY 861 CAATCAAAATTAATCATCATGACTACATTCATTCCTCCATCGCTGAGGACAGAAATTCATT 920
Db 584 CAATCAAAATTAATCATCATGACTACATTCATTCCTCCATCGCTGAGGACAGAAATTCATT 543
QY 921 GTGTGGCATTTGATTTGATGTCAGCTCTATTCAATACCTTCTCTCTCAGATGATGTA 980
Db 644 GTGTGGCATTTGATTTGATGTCAGCTCTATTGAATACCTTCTCTCTCAGATGATGTA 703
QY 981 AGATCAAAAGAAATTCGAAGGAGTTGTAACCGCTGTGTGTACATGCTTTGCTCA 1040
Db 704 AGATCAAAAGAAATTCGAAGGAGTTGTAACCGCTGTGTGTACATGCTTTGCTCA 763
QY 1041 CTCATGTGGATAGCATGGATTTGATTACAAAGGTCACCTTATAGAAATAGAGATGTG 1100
Db 764 CTCATGTGGATAGCATGGATTTGATTACAAAGGTCACCTTATAGAAATAGAGATGTG 823
QY 1101 AGCTGTGAGGTCACAGCTAGAGGAGTCCAAAGAAATCTGGATTTGCTTTCTGACA 1160
Db 824 TGCTGTGAGGTCACAGCTAGAGGAGTCCAAAGAAATCTGGATTTGCTTTCTGACA 883
QY 1161 TCTCGTGTGTAGCAATTTCTCTCTGAGTGGAGCTGGAACCTGTAAAGGATGTCTAA 1220
Db 884 TCTCGTGTGTAGCAATTTCTCTCTGAGTGGAGCTGGAACCTGTAAAGGATGTCTAA 943
QY 1221 TTCTTTCTGCTGAGACGAAATGCTATGGCTCAGATGACTTCTTAGAGGATTTGCCCT 1280
Db 944 TTCTTTCTGCTGAGACGAAATGCTATGGCTCAGATGACTTCTTAGAGGATTTGCCCT 1003
QY 1281 TTGAGCAAAATAGGTA 1295
Db 1004 TTGAGCAAAATAGGTA 1018
RESULT 8
E01791
LOCUS
DEFINITION DNA sequence coding for non-A non-B hepatitis specific antigen
protein.
E01791
ACCESSION
VERSION E01791.1 GI:2170044
KEYWORDS JP 1989002576-A/1.
SOURCE Pan paniscus (pygmy chimpanzee)
ORGANISM Pan paniscus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Takahashi, K., Shibui, T., Uchida, M., Kurebayashi, R., Teranishi, Y.,
Nakanishi, S. and Kitamura, N.
DNA FRAGMENT
Patent: JP 1989002576-A 1 06-JAN-1989;
MITSUBISHI KASEI CORP
COMMENT OS Pan paniscus (Chimpanzee)
PN JP 1989002576-A/1
PD 06-JAN-1989

PF 04-JUN-1987 JP 1987140586
PR 31-MAR-1987 JP 87P 78313
PI TAKAHASHI KAZUNOBU, SHIBUI TATSURO, UCHIDA MICHIRU, PI
KUREBAYASHI RIE,
PI TERANISHI YUTAKA, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
C12N15/00, (C12N15/00, C12R1:91);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue type=Liver;
FH Key Location/Qualifiers
FT mat_peptide 1..1332
FT antigen protein.
FT /product="Non-A non-B hepatitis specific
FT Location/Qualifiers
1..1332
/organism="Pan paniscus"
/mol_type="genomic DNA"
/db_xref="taxon:9597"
ORIGIN
Query Match 18.1%; Score 448.6; DB 6; Length 1332;
Best Local Similarity 94.1%; Pred. No. 2.2e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 801 AATATATTTTGAATTAATATCTTGTCTTATGCTACCTTACAGTTAATCCATGGAA 860
Db 797 ACATATCCTACATCTTGAACGGTAACATTCGTGTAGATACCAAGTTAATCCATGGAA 856
QY 861 CAATCAAAATTAATCATCATGACTACATTCATTCCTCCATCGCTGAGGACAGAAATTCATT 920
Db 857 CAATCAAAATTAATCATCATGACTACATTCATTCCTCCATCGCTGAGGACAGAAATTCATT 916
QY 921 GTGTGGCATTTGATTTGATGTCAGCTCTATTCAATACCTTCTCTCTCAGATGATGTA 980
Db 917 GTGTGGCATTTGATTTGATGTCAGCTCTATTGAATACCTTCTCTCTCAGATGATGTA 976
QY 981 AGATCAAAAGAAATTCGAAGGAGTTGTAACCGCTGTGTGTACATGCTTTGCTCA 1040
Db 977 AGATCAAAAGAAATTCGAAGGAGTTGTAACCGCTGTGTGTACATGCTTTGCTCA 1036
QY 1041 CTCATGTGGATAGCATGGATTTGATTACAAAGGTCACCTTATAGAAATAGAGATGTG 1100
Db 1037 CTCATGTGGATAGCATGGATTTGATTACAAAGGTCACCTTATAGAAATAGAGATGTG 1096
QY 1101 AGCTGTGAGTCCAGCTAGAGGAGTCCAAAGAAATCTGGATTTGCTTTCTGACA 1160
Db 1097 TGCTGTGAGTCCAGCTAGAGGAGTCCAAAGAAATCTGGATTTGCTTTCTGACA 1156
QY 1161 TCTCGTGTGTAGCAATTTCTCTCTGAGTGGAGCTGGAACCTGTAAAGGATGTCTAA 1220
Db 1157 TCTCGTGTGTAGCAATTTCTCTCTGAGTGGAGCTGGAACCTGTAAAGGATGTCTAA 1216
QY 1221 TTCTTTCTGCTGAGACGAAATGCTATGGCTCAGATGACTTCTTAGAGGATTTGCCCT 1280
Db 1217 TTCTTTCTGCTGAGACGAAATGCTATGGCTCAGATGACTTCTTAGAGGATTTGCCCT 1276
QY 1281 TTGAGCAAAATAGGTA 1295
Db 1277 TTGAGCAAAATAGGTA 1291
RESULT 9
E01975
LOCUS
DEFINITION DNA encoding type non-A non-B hepatitis specific antigen.
E01975
ACCESSION
VERSION E01975.1 GI:2170223
KEYWORDS JP 1989124387-A/1.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Ekaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
1 (bases 1 to 1332)
Shibui, T., Kurebayashi, R., Uesono, M., Teranishi, Y., Takahashi, K.,
Nakanishi, S., and Kitamura, N.
MANIFESTATION VECTOR HAVING DNA CODING NON-A NON-B HEPATITIS
SPECIFIC ANTIGEN, TRANSFORMANT AND PRODUCTION OF SAID ANTIGEN
Patent: JP 1989124387-A 1 17-MAY-1989;
MITSUBISHI KASEI CORP
OS Pan troglodytes
PN JP 1989124387-A/1
PD 17-MAY-1989
PF 10-NOV-1987 JP 1987283990
PI SHIBUI TATSURO, KUREBAYASHI RIE, UESONO MICHIRU, PI
TERANISHI YUTAKA,
PI TAKAHASHI KAZUOBU, NAKANISHI SHIGETADA, KITAMURA NAOMI PC
C12N15/00 A61K39/29 C12N1/20 C12P21/02 (C12P21/02, C12R1:19); CC
strandedness: Double;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue types=Liver;
CC *source: clone=PCDVCL-1;
CC Feature is identified by experimental;
CC Key Location/Qualifiers
FH Key
FH mat_peptide 1..1332
FT /product='type non-A non-B hepatitis specific
FT antigen'
FT Location/Qualifiers
source
1..1332
/organism="Pan troglodytes"
/mol_type="genomic RNA"
/db_xref="taxon:9598"

ORIGIN

Query Match 18.1%; Score 448.6; DB 6; Length 1332;
Best Local Similarity 94.1%; Pred. No. 2.1e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 801 AAATATTTTGAATTAATATCTTCTTGTATGCTACCTTACAGTTTAATCCCATGGAAT 860
DB 797 ACATATCTCTACATCTTGAACGGTAACTCGTGATAGATACAGTTTAATCCCATGGAAT 856
QY 861 CAATCAAAATTAATCATCATGACTACTGATTCATTCATCCCATCGCTGAAGACAGAGATTCAAT 920
DB 857 CAATCAAAATTAATCATCATGACTACTGATTCATTCATCCCATCGCTGAAGACAGAGATTCAAT 916
QY 921 GTGTGGCATTTGTATTTGATGCGAGCTCTATTCAATATCTTCTCTCTCAGATGATGATA 980
DB 917 GTGTGGCATTTGTATTTGATGCGAGCTCTATTCAATATCTTCTCTCTCAGATGATGATA 976
QY 981 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGCTGGTGTGTATCATCTGCTTGTCTCA 1040
DB 977 AGATCAAAAGAAATTCGAAGGAGTTGGTAAACGCTGGTGTGTATCATCTGCTTGTCTCA 1036
QY 1041 CTCATGTGATGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1100
DB 1037 CTCATGTGATGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1096
QY 1101 AGCCTGTGAGTCCAGCTAGAGGAGTCCAAAGAAACTTGGATTCTCTTTCTGACA 1160
DB 1097 TGCCTGTGAGTCCAGCTAGAGGAGTCCAAAGAAACTTGGATTCTCTTTCTGACA 1156
QY 1161 TCTCGTGTGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1220
DB 1157 TCTCGTGTGATGATGATTTGATTAACAAAGGTGACCTTATAGAAATAGAGAGATGTG 1216
QY 1221 TCTCTTCTCTCTGAGACCAATGCTATGCTGTGAGTGGAGCTGAGACCTTAAAGGATGTTCTAA 1280
DB 1217 TCTCTTCTCTCTGAGACCAATGCTATGCTGTGAGTGGAGCTGAGACCTTAAAGGATGTTCTAA 1276
QY 1281 TTGAGCAAAATAGGTA 1295

Db 1277 TTGAGCAAAATAGGTA 1291

RESULT 10
CHPP44
LOCUS
DEFINITION
Pan troglodytes gene for non-A non-B hepatitis-associated
microtubular aggregates protein p44, complete cds.
D90034
VERSION
D90034.1 GI:218575
KEYWORDS
Pan troglodytes (chimpanzee)
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE
1 (bases 1 to 1665)
Takahashi, K., Kitamura, N., Shibui, T., Kamizono, M., Matsui, R.,
Yoshiyama, Y., Maeda, T., Kondo, J., Honda, Y., Yamada, E.,
Shimizu, Y.K., Teranishi, Y. and Nakanishi, S.,
Cloning, sequencing and expression in Escherichia coli of cDNA for
a non-A, non-B hepatitis-associated microtubular aggregates protein
J. Gen. Virol. 71 (Pt 9), 2005-2011 (1990)
91011346
PUBLISHED
2170570
COMMENT
These data kindly submitted in computer readable form by:
Yoshikazu Honda
Mitsubishi Kasei Co.
1000, Kamoshida-cho, Midori-ku
Yokohama 227
Japan
Phone: 81-45-963-3454
Fax: 81-45-963-3992.
Location/Qualifiers
FEATURES
source
1..1665
/organism="Pan troglodytes"
/mol_type="mRNA"
/db_xref="taxon:9598"
/clone="PCDVCL-1"
/cell_type="hepatocyte"
/note="CH41"
55..1389
/codon_start=1
/product="microtubular aggregates protein p44"
/protein_id="BAA14082.1"
/db_xref="GI:218576"
/translation="MAVTRLTWLEKILQNHFGKRLSLYLKSGVGHNGVLLDRC
CNGPTLVITYSEDHIGAVAEQYKVASIILFALQETKISWKLGLYPTLFC
CDVAKYNTPTQIDGRNKKVIMDLAKTNELGACNCTISIDYEFVCEDESIDRRI
KGVIEURKSLALRTYEPYISLQVQIRILLGLPIGAKSSFFNSVSVFQGHVTHQA
LVGNTPTGISSEKTYTISIDKQKYLPIELCDLSGLSEKGLGCHDDISYLNGIR
DRYQNPMSIKLNHHDYIDPSLDKIDIERCPVRSKLEEVQKLFALSDISVVSNY
VNAGVVHVALTHVDSMDLITKGLDIERCPVRSKLEEVQKLFALSDISVVSNY
SSEWELDPVKDVLILSALRMLWAADDFLEDLFPFEIQNLNREIINCAQGGK"

misc_feature 466..474
/note="put. N-glycosylation site"
misc_feature 721..729
/note="put. N-glycosylation site"
misc_feature 1225..1233
/note="put. N-glycosylation site"
polyA_signal 1628..1633
ORIGIN

Query Match 18.1%; Score 448.6; DB 9; Length 1665;
Best Local Similarity 94.1%; Pred. No. 2.1e-78;
Matches 466; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 801 AAATATTTTGAATTAATATCTTCTTGTATGCTACCTTACAGTTTAATCCCATGGAAT 860
DB 851 ACATATCTCTACATCTTGAACGGTAACTCGTGATAGATACAGTTTAATCCCATGGAAT 910
QY 861 CAATCAAAATTAATCATCATGACTACTGATTCATTCATCCCATCGCTGAAGACAGAGATTCAAT 920


```

CC *source: tissue type=Liver;
CC *source: clone=CDVCL-1;
CC Feature is identified by experimental;
FH Key
FH Location/Qualifiers
FT misc_feature 1..1667
FT /note=DNA containing complete sequence of FT
FT type nonA nonB
FT Location/Qualifiers
FT 1..1667
FT /organism="Pan troglodytes"
FT /mol_type="genomic RNA"
FT /db_xref="taxon:9598"

FEATURES
  source
  ORIGIN
    Query Match 18.1%; Score 448.6; DB 6; Length 1667;
    Best Local Similarity 94.1%; Pred. No. 2.1e-78; Indels 0; Gaps 0;
    Matches 466; Conservative 0; Mismatches 23;
  QY 801 AAATATTTTGAATTAATATCTTCTGCTTTATGTCTACCTTACAGTTTAAATCCCATGGAAT 860
  DB 853 ACATATCTTACATCTTGAACGGTAACATTCGGTATAGATACCAAGTTTAAATCCCATGGAAT 912
  QY 861 CAATCAAAATTAATCATCATGACTACATGATTCCTCCATCCGTCGAGGACAGAAATTCATT 920
  DB 913 CAATCAAAATTAATCATCATGACTACATGATTCCTCCATCCGTCGAGGACAGAAATTCATT 972
  QY 921 GTGTGGCAATTTGTATTTGATGCGAGCTCTATTCAATACTTCTCTCTCAGATGATAGTAA 980
  DB 973 GTGTGGCAATTTGTATTTGATGCGAGCTCTATTGAATACTTCTCTCTCAGATGATAGTAA 1032
  QY 981 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAAGCTGGTGGTACATGCTGGCTTTGCTCA 1040
  DB 1033 AGATCAAAAGAAATTCGAAGGGAGTTGGTAAAGCTGGTGGTACATGCTGGCTTTGCTCA 1092
  QY 1041 CTCATGTGATAGCATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGATGCTG 1100
  DB 1093 CTCATGTGATAGCATGATTTGATTACAAAGGTGACCTTATAGAAATAGAGATGCTG 1152
  QY 1101 AGCTGTGAGGTCCAGCTAGAGGAAGTCCAAAGAAACTTGGATTTGCTCTTCTGACA 1160
  DB 1153 TGCTGTGAGGTCCAGCTAGAGGAAGTCCAAAGAAACTTGGATTTGCTCTTCTGACA 1212
  QY 1161 TCTCGTGGTTAGCAATTAATCTCTGAGTGGAGCTGGACCTCTAAAGGATGTTCTAA 1220
  DB 1213 TCTCGTGGTTAGCAATTAATCTCTGAGTGGAGCTGGACCTCTAAAGGATGTTCTAA 1272
  QY 1221 TTCTTTCTCTGAGACGAATGCTATGGCTGCGAGATGACTTCTTAGAGGATTTGCTTT 1280
  DB 1273 TTCTTTCTCTGAGACGAATGCTATGGCTGCGAGATGACTTCTTAGAGGATTTGCTTT 1332
  QY 1281 TTGAGCAAAATAGTA 1295
  DB 1333 TTGAGCAAAATAGGA 1347

RESULT 13
AX284603
LOCUS
DEFINITION Homo sapiens 567 bp DNA linear PAT 20-NOV-2001
ACCESSION Sequence 408 from Patent WO0179556.
VERSION AX284603
KEYWORDS AX284603.1 GI:17045291
SOURCE
  ORGANISM
    Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Lillie, J., Brown, J.L., Bolt, A. and van Huffel, C.
  Novel genes, compositions and methods for the identification,
  assessment, prevention, and therapy of human cancers
  Patent: WO 0179556-A 408 25-OCT-2001;

REFERENCE
  1
  TITLE
  JOURNAL

```

```

Millennium Predictive Medicine, Inc. (US)
  Location/Qualifiers
  1..567
  /organism="Homo sapiens"
  /mol_type="unassigned DNA"
  /db_xref="taxon:9606"

ORIGIN
  Query Match 16.7%; Score 415.4; DB 6; Length 567;
  Best Local Similarity 98.4%; Pred. No. 8.6e-72;
  Matches 419; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
  QY 1116 AGCTAGAGAGTCCAAAGAAACTTGGATTTGCTTTCTGACATCTCGGTGGTTAGCA 1175
  DB 142 AGCTAGAGAGTCCAAAGAAACTTGGATTTGCTTTCTGACATCTCGGTGGTTAGCA 201
  QY 1176 ATTATTTCTCTGAGTGGGAGCTGGACCCCTGTAAAGGATCTTCTAATTTCTTCTGCTGA 1235
  DB 202 ATTATTTCTCTGAGTGGGAGCTGGACCCCTGTAAAGGATCTTCTAATTTCTTCTGCTGA 261
  QY 1236 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTTTGAGCAATAGGTA 1295
  DB 262 GACGAATGCTATGGGCTGCAGATGACTTCTTAGAGGATTTGCTTTTGAGCAATAGGTA 321
  QY 1296 GATGCTTTGGTGGTGGAGGCTTGAAGCGGTCAAGTAGTTGGCTACTTTCTGCTTGA 1355
  DB 322 GATGCTTTGGTGGTGGAGGCTTGAAGCGGTCAAGTAGTTGGCTACTTTCTGCTTGA 381
  QY 1356 TCATTAATAATACCTGGCAGCTCTCTGCTTTTGTGGGTGTGCGCTGATGATTCTTCT 1415
  DB 382 TCATTAATAATACCTGGCAGCTCTCTGCTTTTGTGGGTGTGCGCTGATGATTCTTCT 441
  QY 1416 GCATTTTAAACCCACTCCCTGGATGCATTTTCCCTCTTTCGATTTCCCTCTTTCTCGGA 1475
  DB 442 GCATTTTAAACCCACTCCCTGGATGCATTTTCCCTCTTTCGATTTCCCTCTTTCTCGGA 501
  QY 1476 GTTCATACATAGAGATCTGCACATGATGTTTTCCTTTTGTCTTGTAGAGTAAAGTTTAA 1535
  DB 502 AGTCATACATAGAGATCTGCACATGATGTTTTCCTTTTGTCTTGTAGAGTAAAGTTTAA 561
  QY 1536 AATAAT 1541
  DB 562 AATAAT 567

RESULT 14
D28908S8
LOCUS
DEFINITION Homo sapiens gene for hepatitis C-associated microtubular aggregate
  protein p44, exon 9 and complete cds.
ACCESSION D28915
VERSION D28915.1 GI:598954
KEYWORDS
  SEGMENT
  SOURCE
    ORGANISM
      Homo sapiens (human)
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      Kitamura, A., Takahashi, K., Okajima, A. and Kitamura, N.
      Induction of the human gene for p44, a hepatitis-C-associated
      microtubular aggregate protein, by interferon-alpha/beta
      Eur. J. Biochem. 224 (3), 877-883 (1994)
      95010078
      MEDLINE
      PUBMED 7925411
      REFERENCE
        2 (bases 1 to 483)
        Kitamura, N.
        Direct Submission
        Submitted (11-MAR-1994) Naomi Kitamura, Institute for Liver
        Research, Kansai Medical University, Moriguchi, Osaka 570, Japan
        (tel:06-992-1001(ex.2530), Fax:06-994-6095)
        Location/Qualifiers
        1..483
      JOURNAL

```

TITLE

NOVEL GENES, COMPOSITIONS AND METHODS FOR THE IDENTIFICATION,

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:00:48 ; Search time 2810.53 Seconds
(without alignments)
16458.288 Million cell updates/sec

Title: US-10-051-835-17
Perfect score: 1549
Sequence: 1 aacacacagatatttcag.....ctctaaatctttaccatcaa 1549

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	57.4	1200	13	BX447355
2	814.2	52.6	1062	12	EM471130
3	617.2	39.8	972	10	B8888763
4	614.4	39.7	616	9	AL603127

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	601	38.8	601	14	CB134607
6	570.6	36.8	611	13	BX473904
7	538	34.7	538	14	CB123846
8	531	34.3	960	13	BX457116
9	489.6	31.6	3953	11	AK052535
10	433.6	28.0	762	9	AL701060
11	415.4	26.8	768	12	BG615970
12	377.6	24.4	602	14	CD706751
13	368.2	23.8	541	10	AW532890
14	358.6	23.2	664	9	AL700018
15	351.8	22.7	572	13	BX473839
16	335.6	21.7	553	10	B5683286
17	330	21.3	545	14	CB105270
18	326.6	21.1	624	9	AL700011
19	326.6	21.1	624	9	AL7000902
20	320.8	20.7	440	10	B562499
21	309	19.9	839	29	CC523935
22	273	17.6	308	9	AI370412
23	267.6	17.3	618	28	AZ383370
24	260.2	16.8	705	14	CF745483
25	246.2	15.9	379	10	AW526795
26	239.4	15.5	361	10	B5097316
27	190.8	12.3	739	28	BZ206599
28	118.6	7.7	182	10	B562500
29	117	7.6	117	12	BG272761
30	101	6.5	785	28	BZ608978
31	93.2	6.0	220	10	AW260345
32	80.6	5.8	786	14	CK018177
33	88.2	5.7	1811	29	CG753732
34	84.2	5.4	408	14	CB770357
35	77.8	5.0	295	12	BG942089
36	74.8	4.8	1124	13	BX436282
37	73.8	4.8	1566	29	CG757757
38	73.6	4.8	495	29	BX217750
39	73.4	4.7	1531	29	CG748014
40	73	4.7	187	14	CD701571
41	71.6	4.6	342	29	AY406818
42	71.6	4.6	437	9	AI087313
43	71.6	4.6	453	10	BEG75526
44	71.6	4.6	467	9	AI457736
45	71.6	4.6	499	9	AI827737

ALIGNMENTS

RESULT 1
BX447355
LOCUS
DEFINITION
S-PRIME, mRNA sequence.
ACCESSION
BX447355
VERSION
BX447355.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM

BX447355 1200 bp mRNA linear EST 22-MAY-2003
BX447355 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP006YK08

REFERENCE
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 151 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4292.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AAM13ZH08QPl&cluster=4292.r>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

CB134607 K-EST0186
BX473904 DKF2686P
CB123846 K-EST0172
BX457116 BX457116
AK052535 Mus muscu
AL701060 DKF2686T
BG615970 60243421
CD706751 EST3278
AW532890 UI-R-BUO-
AL700018 DKF2686C
BX473839 DKF2686H
B5683286 182767 MA
CB105270 K-EST0112
AL700011 DKF2686N
AL7000902 DKF2686N
BF562499 UI-R-BUO-
CC523935 CH240_372
AI370412 qv77d02.x
AZ383370 1M0141D04
CF745483 UI-M-GVO-
AW526795 UI-R-B01-
BE097316 UI-R-B01-
BZ206599 CH240-436
BF562500 UI-R-BUO-
BG272761 nah89d09.
BZ608978 WHAD05TR
AW260345 um77b01.x
CK018177 AGENCOURT
CG753732 P048-4-G0
CB770357 AMGNNUC:N
BG942089 ax21a01.x
BX436282 BX436282
CG757757 P053-1-D0
BX217750 Danilo ter
CG748014 P041-4-B0
CD701571 EST18095
AY406818 Homo sapi
AI087313 oz45e10.x
BEG75526 7f10g07.x
AI457736 tj49a01.x
AI827737 wf11e07.x

Faraday Avenue Genoscope sequence ID : CS0AAW13ZH08QPI.

FEATURES

source

1. .1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP006YK08"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 57.4%; Score 889; DB 13; Length 1200;
Best Local Similarity 99.3%; Pred. No. 3.9e-149;
Matches 903; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
QY 68 AGACTACTGGGTATCAGATGGCAAGCAACTTAACCTCCAGAAATACAGATGTTGG 127
DB 282 AGTCAGTTTGGTATCAGATGGCAAGCAACTTAACCTCCAGAAATACAGATGTTGG 341
QY 128 ACAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAA 187
DB 342 ACAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAA 401
QY 188 ATTTTGACTGAACAAGACGTAATGGAGCAGTCTTGAAGTGGTTAAAAAAGNACATCTT 247
DB 402 ATTTTGACTGAACAAGACGTAATGGAGCAGTCTTGAAGTGGTTAAAAAAGNACATCTT 461
QY 248 GTTGATATGGCATCACATGACACGATGATTAATCAATAGAAAGTATTAATCAAGAAATG 307
DB 462 GTTGATATGGCATCACATGACACGATGATTAATCAATAGAAAGTATTAATCAAGAAATG 521
QY 308 CGGAAAACAGCCATTTGAAGATTCGATTCAGACATCTTAAGATGGGAAGCCAGTAAATAT 367
DB 522 CGGAAAACAGCCATTTGAAGATTCGATTCAGACATCTTAAGATGGGAAGCCAGTAAATAT 581
QY 368 GCTCCTTAAGACCAAACTGTGTCTCAAAAGGAAGCTAGAGAACTTCAAGAACAAAACAA 427
DB 582 GCTCCTTAAGACCAAACTGTGTCTCAAAAGGAAGCTAGAGAACTTCAAGAACAAAACAA 641
QY 428 AAGGTTAAGAGAACCCAGATATGGCTTAATCCGTTGCAATAGTACAACTGTCTAAAGGT 487
DB 642 AAGGTTAAGAGAACCCAGATATGGCTTAATCCGTTGCAATAGTACAACTGTCTAAAGGT 701
QY 488 TCTAAGTCACTAAAGTTGAGCTCATAGAGATAAATAGATTATACAAAGGAAGGCAG 547
DB 702 TCTAAGTCACTAAAGTTGAGCTCATAGAGATAAATAGATTATACAAAGGAAGGCAG 761
QY 548 CCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGT 607
DB 762 CCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGT 821
QY 608 TACAGTTGGATTTTGTCTACAGCTTGAACAGGACCGGCAATCTCATTTGATCCGATA 667
DB 822 TACAGTTGGATTTTGTCTACAGCTTGAACAGGACCGGCAATCTCATTTGATCCGATA 881
QY 668 CATGAATTCAAAGCCTTCAAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAA 727
DB 882 CATGAATTCAAAGCCTTCAAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAA 941
QY 728 TTTAGCAATGAGGTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGGCACT 787
DB 942 TTTAGCAATGAGGTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGGCACT 1001
QY 788 ATTCAATTTGGAGTCAAAAGCAAAACCCCATGGGAAAAATTTGTTGGCATCAAAAGTCACCAAT 847
DB 1002 ATTCAATTTGGAGTCAAAAGCAAAACCCCATGGGAAAAATTTGTTGGCATCAAAAGTCACCAAT 1061
QY 848 GATACCAAGGAAGCCCTCATTAACCAATTTCAATCTGATGATGAATTAACCAAGATTTTGAAGAC 907

DB 1062 GATACCAAGGAAGCCCTCATTAACCAATTTCAATCTGATGATAAACAGTATTTGAAGAC 1121
QY 908 CATCAAGTCCAAACAAAGCAAGAGAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTG 967
DB 1122 CATCAAGTCCAAACAAAGCAAGAGAGTGCATTCGAGAGCCAGATTTGTGGAAG-TTACTG 1180
QY 968 CCAATAGT 976
DB 1181 CCAATAGT 1189
RESULT 2
LOCUS BM4711130 1062 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6463894 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5517666
5', mRNA sequence.
ACCESSION BM4711130
VERSION BM4711130.1 GI:18520172
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1062)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: csapb-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUAM12177 row: d column: 19
High quality sequence stop: 667.
Location/Qualifiers
1. .1062
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5517666"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
ORIGIN
Query Match 52.6%; Score 814.2; DB 12; Length 1062;
Best Local Similarity 94.9%; Pred. No. 9.1e-136;
Matches 897; Conservative 0; Mismatches 38; Indels 10; Gaps 5;
QY 187 AATTTTGATGACCAAGACGTTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCT 246
DB 1 AATTTTGATGACCAAGACGTTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCT 60
QY 247 TGTTCATATGGCATCACATGGAGCCAGCTATTCAATAGAGAACTTTCAAGATTT 306
DB 61 TGTTCATATGGCATCACATGGAGCCAGCTATTCAATAGAGAACTTTCAAGATTT 120
QY 307 GCGGAAAACAGCCATTGAAGATTCGATTCAGACATCTAAGATGGAAAGCCAGTAAAAA 366
DB 121 GCGGAAAACAGCCATTGAAGATTCGATTCAGACATCTTGAATGGAAAGCCAGTAAAAA 180
QY 367 TGCTCCTTAAGACCAAACTGTGTCTCAAAAGGAAGCGTAGAGAACTTCAAGAGCAAAACA 426
DB 181 TGCTCCTTAAGACCAAACTGTGTCTCAAAAGGAAGCGTAGAGAACTTCAAGAGCAAAACA 240


```
QY 427 AAAGGTAAGAGAACCCAGATATGGCTTAATCCGCTGCAATGAGTACAACTGCTAAAGG 486
Db 241 AAAGGTAAGAGAACCCAGATATGGCTTAATCCGCTGCAATGAGTACAACTGCTAAAGG 300
QY 487 TTCTAAGTCACATTAAGTTGAGCTCATAGAGATATAAATAGATTATACAAAGGAAGGCA 546
Db 301 TTCTAAGTCACATTAAGTTGAGCTCATAGAGATATAAATAGATTATACAAAGGAAGGCA 360
QY 547 ACCATCCATAGACCTGACATGTGTATCATATCCATTTTGATGAATTCAGTAAATCCATATCG 606
Db 361 ACCATCCATAGACCTGACATGTGTATCATATCCATTTTGATGAATTCAGTAAATCCATATCG 420
QY 607 TTACAGATTGATTTTGTAGTCTACAGCTTGAACAGGACCGCAATCTCATTTGATTCGGAT 666
Db 421 TTACAGATTGATTTTGTAGTCTACAGCTTGAACAGGACCGCAATCTCATTTGATTCGGAT 480
QY 667 ACATGAATTCAGAGCTTTCACAAATACAGCAACGCCACAGAGAGGATGTCAAGATGAA 726
Db 481 ACATGAATTCAGAGCTTTCACAAATACAGCAACGCCACAGAGAGGATGTCAAGATGAA 540
QY 727 ATTAGCAATGAGTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGGCAC 786
Db 541 ATTAGCAATGAGTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGGCAC 600
QY 787 TATTCATTTTGGAGTCAAGACAAACCCCATGGGAAATTTGTGGCATCAAAGTCACCAA 846
Db 601 TATTCATTTTGGAGTCAAGACAAACCCCATGGGAAATTTGTGGCATCAAAGTCACCAA 660
QY 847 TGATACCAAGAGCCCTCATTAACCATTCATCTGATGATGAATTAACAGATTTTGAAGA 906
Db 661 TGATACCAAGAGCCCTCATTAACCATTCATCTGATGATGAATTAACAGATTTTGAAGA 720
QY 907 CCATCAAGTCCCAACAGCAAGAGAGTGCATTCGAGAGCAAGATTTGTGGAGTTTACT 966
Db 721 CCATCAAGTCCCAACAGCAAGAGAGTGCATTCGAGAGCAAGATTTGTGGAGTTTACT 780
QY 967 GCCAATAGTACTTATCTGA-CAGATTGTTATG-AAGTGGACATTTCCAGTTTC 1024
Db 781 GCCAATAGTACTTATCTGA-CAGATTGTTATG-AAGTGGAGTGCATTTATCCACCGTTC 840
QY 1025 T-CTGATGCAATATGATPAC-TTCCAGATTAATGCAAAATTT-ACAACACA 1076
Db 841 TTCTGAAGGCGCAATATGATTAATTTCCGATTAATGCGCAATTTCCCGCCACCCCA 900
QY 1077 AAATATGGGAACAAAGTAAATAATTTCTCACTATTTTGTGCGAGATG 1121
Db 901 TTATGGGAACCAAGGTAAATAATTTCTCCCTTATTTGGGCGAAG 945

RESULT 3
BE888763
LOCUS
DEFINITION BE888763 972 bp mRNA linear EST 20-OCT-2000
601513677F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914995 5',
mRNA sequence.
ACCESSION BE888763
VERSION BE888763.1 GI:10345391
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 972)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
```

```
http://image.llnl.gov
Plate: ILAM9737 row: j column: 20
High quality sequence stop: 637.

FEATURES
            Location/Qualifiers
            1..972
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3914995"
            /tissue_type="leiomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 71"
            /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2.1 Kb."

ORIGIN
Query Match      39.8%; Score 617.2; DB 10; Length 972;
Best Local Similarity 89.3%; Pred. No. 1.4e-100;
Matches 725; Conservative 0; Mismatches 73; Indels 14; Gaps 5;

QY 447 ATATGGCTAATCG-GTCTGCAATGAGTACAACTGCTAAAGTTCTTAAGTCACCTAAAGTT 505
Db 1 ATATGGCTAATCGGTTCTGCAATGAGTACAACTGCTAAAGTTCTTAAGTCACCTAAAGTT 60
QY 506 GAGCTCATAGAGATAAAATAGATTATACAAAGGAAGCAACCATCCATAGACCTGACA 565
Db 61 GAGCTCATAGAGATAAAATAGATTATACAAAGGAAGCAACCATCCATAGACCTGACA 120
QY 566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGATTTAGT 625
Db 121 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGATTTAGT 180
QY 626 CTACAGCTTGAACAGGACGAGCAATCTCATTTGATCCGATACATGAATTCAAAGCCTTC 685
Db 181 CTACAGCTTGAACAGGACGAGCAATCTCATTTGATCCGATACATGAATTCAAAGCCTTC 240
QY 686 ACAATACAGCAACAGCCACAGAGAGATGTCAAGATGAATTTAGCAATGAGGTTTC 745
Db 241 ACAATACAGCAACAGCCACAGAGAGATGTCAAGATGAATTTAGCAATGAGGTTTC 300
QY 746 CGATTTGCTTCAGTTGTATGAATTCACGTACCAATGGCACTATTCTATTTGGAGTCAA 805
Db 301 CGATTTGCTTCAGTTGTATGAATTCACGTACCAATGGCACTATTCTATTTGGAGTCAA 360
QY 806 GACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCAACATGATACCAAGGAGCCCTC 865
Db 361 GACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCAACATGATACCAAGGAGCCCTC 420
QY 866 ATTACGATTTCAATCTGATGATACCAAGTATTTTGAAGACCATCAAGTCCAAACAGCA 925
Db 421 ATTACGATTTCAATCTGATGATACCAAGTATTTTGAAGACCATCAAGTCCAAACAGCA 480
QY 926 AAGAAGTGCATTCGAGAGCCAAAGATTTGTGGAAGTTTACTGCCAAATAGTACTTATCT 985
Db 481 AAGAAGTGCATTCGAGAGCCAAAGATTTGTGGAAGTTTACTGCCAAATAGTACTTATCT 540
QY 986 GACAGATTTGTT-ATTGAGTGCACATTTATCCAGATTTCTGATGATGATGATGATGAT 1044
Db 541 GACAGATTTGTT-ATTGAGTGCACATTTATCCAGATTTCTCTGGAATGCAATGATGAT 600
QY 1045 CTTCAGATTTAAATGCAAAATTTACAAACCAAAATATGCGG-----AACAAAGTAAAAA 1099
Db 601 TTTCAGATTTAAATGCAAAATTTACAAACCAAAATATGCGGAAACCAAAAGTAAAAACA 660
QY 1100 TTCTCACTATTTGCGGAG---ATGGGACAGCTCTTAAGGACATTAACGAAA---ATGA 1152
Db 661 TTCTCACTATTTGTTGCGGAGATGGGAGTGGGACGCTTACGAGACATTTCCGCAAGCTAA 720
QY 1153 AGTTGATTTTCAGAGCATTTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAGCAG 1212
Db 721 GTGACCTTCCGAGCCTTACAGGCCCGATTTCTTAAACACTTGGCAATCCGAAACAGACG 780
```

```

QY      1213 AGAAGAAAATTCAGAGCAAAACAAATATAA 1244
Db      781 GACAAAATCTCCGAGAGAAATTAACAGA 812

RESULT 4
AL603127
LOCUS      616 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFP686D092 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION  AL603127
VERSION     AL603127.1 GI:15166633
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 616)
AUTHORS     Wambutt,N., Heubner,D., Mewes,W., Weil,B. and Wiemann,S.
TITLE       EST (Wambutt,N., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL     Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No sl sequence available.
            This clone (DKFP686D092) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES             source
     source
     1..616
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="DKFP686D092"
         /dev_stage="adult"
         /lab_host="DH10B"
         /clone_lib="686 (synonym: hlcc3)"
         /note="vector: priplEX2; Site_1: SfIIA; Site_2: SfiIB;
         cDNA-collection"

ORIGIN
Query Match      39.7%; Score 614.4; DB 9; Length 616;
Best Local Similarity 99.8%; Pred. No. 5e-100;
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      140 GTAAATCAGTGGTTAGAAAGCTATPAGATTGACCAAAAACACAGGGAAATTTGACTGAA 199
Db      1   GTAAATCCGTGGTTAGAAAGCTATPAGATTGACCAAAAACACAGGGAAATTTGACTGAA 60

QY      200 CAAGACGTGANTGAGCAGCTTCAAGTGGTTAAAAAGAAACATCTTTGTGATATGGGC 259
Db      61 CAAGACGTGAATGAGCAGCTTCAAGTGGTTAAAAAGAAACATCTTTGTGATATGGGC 120

QY      260 ATCACATGACACAGCTATTCAATAGAGAACTATTCAAGAAATTCGGAAAAACAGCC 319
Db      121 ATCACATGACACAGCTATTCAATAGAGAACTATTCAAGAAATTCGGAAAAACAGCC 180

QY      320 ATTGAAGATTCGATTTCAGACATCTAAGTGGGAAGCCCACTAAATGCTCTCTTAAGAC 379
Db      181 ATTGAAGATTCGATTTCAGACATCTAAGTGGGAAGCCCACTAAATGCTCTCTTAAGAC 240

QY      380 CAAACTGTGTCTCAAAAGGAAACGTAGAGAACTTCAAGCAAAAACAAAAGGGTAAAGAG 439
Db      241 CAAACTGTGTCTCAAAAGGAAACGTAGAGAACTTCAAGCAAAAACAAAAGGGTAAAGAG 300

QY      440 AACCAGATATGGCTAATCCCTCTGCATGATGATCAACTGCTAAAGCTTCTAGTCACTA 499
Db      301 AACCAGATATGGCTAATCCCTCTGCATGATGATCAACTGCTAAAGCTTCTAGTCACTA 360

```

```

QY      500 AAAGTTGAGTCTATAGAGATAAAATAGATTATATAAAGGAAAGGCAACCATCATAGAC 559
Db      361 AAAGTTGAGTCTATAGAGATAAAATAGATTATATAAAGGAAAGGCAACCATCATAGAC 420

QY      560 CTGACATGTGTATCATATCCATTGTAATTCAGTAATCCATATCGTTACAAGTTGGAT 619
Db      421 CTGACATGTGTATCATATCCATTGTAATTCAGTAATCCATATCGTTACAAGTTGGAT 480

QY      620 TTTAGTCTTACAGCCTGAAACAGGACCAAGCAATCTCATTTGATCCGATACATGAATTCAAA 679
Db      481 TTTAGTCTTACAGCCTGAAACAGGACCAAGCAATCTCATTTGATCCGATACATGAATTCAAA 540

QY      680 GCCTTCACAAATACAGCAACAGCCACAGAGAGATGTCAGATGAATTTGCAATGAG 739
Db      541 GCCTTCACAAATACAGCAACAGCCACAGAGAGATGTCAGATGAATTTGCAATGAG 600

QY      740 GTTTTCCGATTTGCTT 755
Db      601 GTTTTCCGATTTGCTT 616

RESULT 5
CB134607
LOCUS      601 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0186079 L14ChoICK0 Homo sapiens cDNA clone L14ChoICK0-13-G04
            5' mRNA sequence.
ACCESSION  CB134607
VERSION     CB134607.1 GI:28100932
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 601)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Ch.K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 13 row: G column: 04
            High quality sequence stop: 601.
            Location/Qualifiers
                1..601
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="L14ChoICK0-13-G04"
                    /sex="M"
                    /cell_line="Choi-CK"
                    /lab_host="Top10P"
                    /clone_lib="L14ChoICK0"
                    /note="Organ: Liver; Vector: p7713-Pac; Site_1: EcoRI;
                    Site_2: NotI; The library was contributed by the Soares
                    laboratory and it was constructed as described by Bonaldo,
                    M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                    6(9): 791-806. RNA was prepared from harvested cell
                    culture."

ORIGIN
Query Match      38.8%; Score 601; DB 14; Length 601;
Best Local Similarity 100.0%; Pred. No. 1.2e-97;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      87 TGGCAAGCAACTTAACTTCAGAAAATACAGATGATTGACAAAAGAGGATGTAAATC 146

```

Db 1 TGGCAAGCAACTTAACCTTCCAGAAATACAGATGATTGGACAAAAGAGGATGTAAATC 60
QY 147 AGTGGTTAGAAAGTCATTAAGATTGACCAAAACACAGAGGAATTTTGACTCAACAGACG 206
Db 61 RTGGTTTGAAGAAGTCATTAAGATTGACCAAAACACAGAGGAATTTTGACTCAACAGACG 120
QY 207 TGAATGGAGCAGCTCTTGAAGTGGTTTAAAAAAGAAACATCTTGTGATATGGGCATCAC 266
Db 121 TGAATGGAGCAGCTCTTGAAGTGGTTTAAAAAAGAAACATCTTGTGATATGGGCATCAC 180
QY 267 ATGGACCACTATTCAATAGAGAACTATTCAAGAAATTCGGGAAACAGCCATTGAG 326
Db 181 ATGGACCACTATTCAATAGAGAACTATTCAAGAAATTCGGGAAACAGCCATTGAG 240
QY 327 ATTCCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAATAATGCTCTTAAAGACCAAACTG 386
Db 241 ATTCCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAATAATGCTCTTAAAGACCAAACTG 300
QY 387 TGTCTCAAAAGAAAGTAGAGAACTTCAAGACCAAAACAAAGGTTAAAGAGAACCCAG 446
Db 301 TGTCTCAAAAGAAAGTAGAGAACTTCAAGACCAAAACAAAGGTTAAAGAGAACCCAG 360
QY 447 ATATGGCTAATCCGCTCTGCAATGAGTACAACTGCTTAAAGGTTCTTAAGTCACTTAAAGTTG 506
Db 361 ATATGGCTAATCCGCTCTGCAATGAGTACAACTGCTTAAAGGTTCTTAAGTCACTTAAAGTTG 420
QY 507 AGCTCATAGAGATTAATAGATTATCAAGAGAAAGCAACCATCCATAGACCTGACAT 566
Db 421 AGCTCATAGAGATTAATAGATTATCAAGAGAAAGCAACCATCCATAGACCTGACAT 480
QY 567 GTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTACAAAGTTGGATTTAGTC 626
Db 481 GTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTACAAAGTTGGATTTAGTC 540
QY 627 TACAGCCTGAACAGAGACAGGCAATCTCATTTGATCCGATACATGAATTCAAAGCCTTCA 686
Db 541 TACAGCCTGAACAGAGACAGGCAATCTCATTTGATCCGATACATGAATTCAAAGCCTTCA 600
QY 687 C 687
Db 601 C 601

RESULT 6
BX473904 611 bp mRNA linear EST 04-SEP-2003
LOCUS DXF2p686P01165 r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DXF2p686P01165 5', mRNA sequence.

ACCESSION BX473904
VERSION BX473904.1 GI:31668147
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Bahr, A., Lauber, J., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Fobo, G., Han, M. and Wiemann, S.
TITLE EST (Bahr, A., Lauber, J., Mewes, H.W., Weil, B., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No 5' sequence available.
This clone (DXF2p686P01165) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

source 1. 611
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKF2p686P01165"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/notes="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI; cDNA-collection"
ORIGIN
Query Match 36.8%; Score 570.6; DB 13; Length 611;
Best Local Similarity 99.3%; Pred. No. 3.3e-92;
Matches 573; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AACACATCAGATATTTTTCAGCACTAAAGAGATGGTTTTCCTCCACATATATGTAAAGAA 60
Db 35 AACACATCAGATATTTTTCAGCACTAAAGAGATGGTTTTCCTCCACATATATGTAAAGAA 94
QY 61 ATTTGCAAGACTACTCGGTATCAGAAATGGCAAGCAACTTAACCTTCCAGAAATATCAGA 120
Db 95 ATTTGCAAGACTACTCGGTATCAGAAATGGCAAGCAACTTAACCTTCCAGAAATATCAGA 154
QY 121 TGATTGCAAAAGAGAGATGTTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACA 180
Db 155 TGATTGCAAAAGAGAGATGTTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACA 214
QY 181 CAGGGAATTTTTCACATCAACAGACGTTGATGGAGCAGTCTTGAAGTGGTTAAAAAAGA 240
Db 215 CAGGGAATTTTTCACATCAACAGACGTTGATGGAGCAGTCTTGAAGTGGTTAAAAAAGA 274
QY 241 ACATCTTGTGATATGGGATCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 275 ACATCTTGTGATATGGGATCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
QY 301 AGAATTGCGGAAACAGCCATTCGAAGATTTCAGATTCAGATTCAGATTCAGATTCAGATTC 360
Db 335 AGAATTGCGGAAACAGCCATTCGAAGATTTCAGATTCAGATTCAGATTCAGATTCAGATTC 394
QY 361 TAAATATGCTCTTAAAGACCAAACTGCTCTCAAGAGAACCTAGAGAACTTCAAGCA 420
Db 395 TAAATATGCTCTTAAAGACCAAACTGCTCTCAAGAGAACCTAGAGAACTTCAAGCA 454
QY 421 AAAACAAAAGGGTAAAGAGAACCCAGATATGGCTAATCGCTCTGCAATGAGTCAAACTGC 480
Db 455 AAAACAAAAGGGTAAAGAGAACCCAGATATGGCTAATCGCTCTGCAATGAGTCAAACTGC 514
QY 481 TAAAGGTTCTAAGTCACTAAAGTTGAGCTCATAGAGATATAAGATATAAGATATAAGGA 540
Db 515 TAAAGGTTCTAAGTCACTAAAGTTGAGCTCATAGAGATATAAGATATAAGATATAAGGA 574
QY 541 AAGCAACCATCCATAGACCTGACATGATGATCATAT 577
Db 575 AAGCAACCATCCATAGACCTGACATGATGATCATAT 611

RESULT 7
CB123846 538 bp mRNA linear EST 29-JAN-2003
LOCUS K-EST0172359 L10ChoCK0 Homo sapiens cDNA clone L10ChoCK0-31-D07 5',
DEFINITION mRNA sequence.
ACCESSION CB123846
VERSION CB123846.1 GI:28083454
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 538)
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.B., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.

```

TITLE      21C Frontier Korean EST Project 2001
JOURNAL    Unpublished (2002)
COMMENT    Contact: Kim YS
           Genome Research Center
           Korea Research Institute of Bioscience & Biotechnology
           52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
           Tel.: +82-42-860-4470
           Fax: +82-42-860-4409
           Email: yongsung@mail.kribb.re.kr
           Plate: 31 row: D column: 07
           High quality sequence stop: 538.
FEATURES   Location/Qualifiers
            1..538
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="L10ChcK0-31-D07"
            /sex="M"
            /cell_line="Cho-Ck"
            /lab_host="Top10f"
            /clone_lib="L10ChcK0"
            /notes="Organ: Liver; Vector: pTT73-Pac; Site_1: EcoRI;
            Site_2: NotI; The library was contributed by the Soares
            laboratory and it was constructed as described by Bonaldo,
            M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
            6(9): 791-806. RNA was prepared from harvested cell
            culture."
ORIGIN
Query Match      34.7%; Score 538; DB 14; Length 538;
Best Local Similarity 100.0%; Pred. No. 2.3e-86;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 165 AGATTGACCAAAAACACAGGGAATTTTGACTGAACAAGACGCTGAATGGAGCAGTCTTGA 224
Db 1 AGATTGACCAAAAACACAGGGAATTTTGACTGAACAAGACGCTGAATGGAGCAGTCTTGA 60
Qy 225 AGTGTTAAAAAAGAACATCTTTGTGATATGGGCATCACATGACACAGCTATTCAAA 284
Db 61 AGTGTTAAAAAAGAACATCTTTGTGATATGGGCATCACATGACACAGCTATTCAAA 120
Qy 285 TAGAAGAACTATTCAAGAAATTCGGAAAAACAGCCATTGAAGATTTCGATTTCAGACATCTA 344
Db 121 TAGAAGAACTATTCAAGAAATTCGGAAAAACAGCCATTGAAGATTTCGATTTCAGACATCTA 180
Qy 345 AGATGGAAAGCCAGTAAATATGCTCTTAAGAACCAAACTGTGTCTCAAAAGGAACGTA 404
Db 181 AGATGGAAAGCCAGTAAATATGCTCTTAAGAACCAAACTGTGTCTCAAAAGGAACGTA 240
Qy 405 GAGAACTTCAAGCAAAAACAAAGGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCG 464
Db 241 GAGAACTTCAAGCAAAAACAAAGGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCG 300
Qy 465 CAATGAGTACAACGTCTAAAGGTTCTAAGTCACTATAAAGTTGAGCTCATAGAGATAAAA 524
Db 301 CAATGAGTACAACGTCTAAAGGTTCTAAGTCACTATAAAGTTGAGCTCATAGAGATAAAA 360
Qy 525 TAGATTATACAAAGGAAGCAACCATCCATAGACCTTGACATGTGTATCATATCCATTGG 584
Db 361 TAGATTATACAAAGGAAGCAACCATCCATAGACCTTGACATGTGTATCATATCCATTGG 420
Qy 585 ATGAATTCAGTAATCCATATCGTTTACAAAGTTGGATTTTGTCTACAGCCTTGAAACAGGAC 644
Db 421 ATGAATTCAGTAATCCATATCGTTTACAAAGTTGGATTTTGTCTACAGCCTTGAAACAGGAC 480
Qy 645 CAGCAATCTCATGTATCCGATCATGAATTCGAAGCCTTCACAAATACAGCAACAGC 702
Db 481 CAGCAATCTCATGTATCCGATCATGAATTCGAAGCCTTCACAAATACAGCAACAGC 538
RESULT 8
BX457116
LOCUS
BX457116 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP006YK08
5-PRIME, mRNA sequence.
BX457116 GI:31030860
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 960)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4292.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP006BF04QPL&cluster=4292.r. Contact :
Feng liang Email : fliang@lifetech.com URL : Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP006BF04QPL.
Location/Qualifiers
1..960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP006YK08"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match      34.3%; Score 531; DB 13; Length 960;
Best Local Similarity 93.2%; Pred. No. 3.5e-85;
Matches 573; Conservative 22; Mismatches 16; Indels 4; Gaps 4;
Qy 58 AGACTACTGGTATCAGAAATGGCAAGCAACTTAACTCCAGAAATACAGATGATTGG 127
Db 299 AGDSAGWTTGGTATCAGAAATGGCAAGCAACTTAACTCCAGAAATACAGATGATTGG 358
Qy 128 ACAAAGAGGATGTAATTCAGTGTGTAGAAAGTCATAAGATTGACCAAAAACACAGGAA 187
Db 359 ACAAAGAGGATGTAATTCAGTGTGTAGAAAGTCATAAGATTGACCAAAAACACAGGAA 418
Qy 188 ATTTTGACTGAACAAGCGTGAATGGAGCAGTCTTGAAGTGGTTAAAAA-AAGACATCT 245
Db 419 ATTTTGACTGAACAAGCGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAARRAGAACATCT 478
Qy 247 TGTGTATATGGGCATCACAT-CGACACGCTATTCAAATAGAGAACTATTCAAAGAAT 305
Db 479 TGTGTATATGGGCATCACATGCGACCGCTATTCAAATAGAGAACTATTCAAAGAAT 538
Qy 306 TCGGAAAAACAGCATTTGAAGATTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAAA 365
Db 539 TCGGAAAAACAGCATTTGAAGATTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAAA 598
Qy 366 ATGCTCTCTAAAGACC-AAACCTGTCTCAAAAGGAACGTAGAGAACTTCAAAAGCAAAA 424
Db 599 ATGCTCTCTAAAGACCAGAACTGTCTCAAAAGGAACGTAGAGAACTTCAAAAGCAAAA 658
Qy 425 CAAAGGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCTGCAATGAGTACACTGCTAAA 484
Db 659 CAAAGGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCTGCAATGAGTACACTGCTAAA 718
Qy 485 GGTTCCTAAGTCACTAAAAAGTTGAGCTCATAGAGATAAAATAGATTATACAAAGGAAGG 544
```

```

Db      719 GGDWCTAAGTCTACTAAAGTTGAGCTCTAGAGATAAAATAGATTATDCAAAGGAAAGG 778
QY      545 CAACCATCATAGACTGCATGCTATCATATCCATTTGATGATTCAGTAATCCATAT 604
Db      779 CAACCATCAWAGACTGCATGCTGTATCATWCCATTTGAAGAAWTCAGTAATCCATAT 838
QY      605 CGTTACAAGTTGGATTTAGTCTACAGCTGCTGAAACAGGACCGCAATCTCATTTGATCCG 664
Db      839 CGTTACAAGTTGGATTTAGTCTACAGCTGCTGAAACAGGACCGCAATCTCATTTGATCCG 897
QY      665 ATACATGATTTCAA 679
Db      898 RAAAAAAGAAAAA 912

RESULT 9
LOCUS   AK052535
DEFINITION Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched library, clone: D43004G20 product: weakly similar to CDNA FLJ20073 FIS, CLONE COL02320 (Homo sapiens), full insert sequence.
ACCESSION AK052535
VERSION   AK052535.1 GI:26095242
KEYWORDS  HTC; CAP trapper;
SOURCE    Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL   Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE   Genome Res. 10 (10), 1617-1630 (2000)
PUBMED    20499374
REFERENCE 2
AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL   Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE   Genome Res. 10 (10), 1617-1630 (2000)
PUBMED    20499374
REFERENCE 3
AUTHORS   Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
JOURNAL   RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
MEDLINE   Genome Res. 10 (11), 1757-1771 (2000)
PUBMED    20530913
REFERENCE 4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
JOURNAL   Functional annotation of a full-length mouse cDNA collection
MEDLINE   Nature 409, 685-690 (2001)
PUBMED    11076861
REFERENCE 5
AUTHORS   The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL   Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
MEDLINE   Nature 420, 563-573 (2002)
PUBMED    12000000
REFERENCE 6
AUTHORS   Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

```

```

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES             source
     Location/Qualifiers
         1..3953
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="PANTOM,DB:D43004G20"
             /db_xref="MGI:2422069"
             /db_xref="taxon:10090"
             /clone="D43004G20"
             /tissue_type="lung"
             /clone_lib="RIKEN full-length enriched mouse cDNA library"
             /dev_stage="13 days embryo"
             /notes="putative weakly similar to CDNA FLJ20073 FIS, CLONE COL02320 (Homo sapiens) [SPRR|Q9NXS8, evidence: FASTA, 50.4%ID, 71.7%length, match=1341]"
misc_feature          328..3952
     /notes="putative weakly similar to CDNA FLJ20073 FIS, CLONE COL02320 (Homo sapiens) [SPRR|Q9NXS8, evidence: FASTA, 50.4%ID, 71.7%length, match=1341]"
ORIGIN
Query Match          31.6%; Score 489.6; DB 11; Length 3953;
Best Local Similarity 61.9%; Pred. No. 6.3e-78;
Matches 962; Conservative 0; Mismatches 549; Indels 43; Gaps 10;
QY      15 TTTTCAGCACTAAAGAGAGATGGTTTCCCCACATATATGTAAGAAATTTGCAAGACTAC 74
Db      257 TTTTTCCTTTTAAACGATTTTATCTAGAGAAATTTGAAATTTGGGGAATAAATTT 316
QY      75 TGGGTATCAGATGGGCAACCACTTAACCTCCAGAAATACAGATGATGACCAAAAG 134
Db      317 GAAGCTTGAACATGAGTGACAAAGTAACTCAACCAAAATTTGATCAAGACTGGACCAAG 376
QY      135 AGGATGTAATTCAGTGGTT--AGAAAGTCATAGATTGACCAAAACACAGGGAATTT 191
Db      377 AACATGTAAGAAATGGTTTACTGAGATTTTACATTTGTGAGAAATATGCTCAATCT 436
QY      192 TGACTGAACAGAGCGTGAATGGAGCAGTCTTGAAGTGGTTTAAAGAAACATCTTGTG 251
Db      437 TGTTCAAAAGAGAGCTAACAGGGATGCTCTCAGAGAAATTAACCTGAAGAGGATCTTAGAG 496
QY      252 ATATGGGCATCACATGGACCGACTTATCAATATAGAGAACTATTCAAGAAATTTGGGA 311
Db      497 AAATGGGCTGGCCACGGGGTCCAGCACTTCTGTATAAACCATGTATACAA--CTAA 553
QY      312 AAACAGGCATTGAAGATTCGATTCAGACATCTTAAGATGGGAAAGCCCGATTAAGTCTC 371
Db      554 TTAGTTCCTCTGAAAGTCACAAATCAAGATTCAGAGAAATTAATGATATAAAATCTCTCA 613
QY      372 CTAAGACCAAACTGTCTCTCAAAAGAAAGCTAGAGAACTTCAAGACCAAAACAAAGG 431
Db      614 CAAAGAACAGCAAAACAAAGCCAAAATGAGGAGAGAAATCACTGATTTTATCCACAGTG 673

```

QY	432	GTAAAGAGAAACCCAGATATGGCTAAATCCGTCTGCAATGAGTACAACTGCTAAAGGTTCTA	491
Db	674	ATCATGCTCTCAGGAGACAG-----GGCAGATGAGAAACAAAACCAAGTCTTAGC	726
QY	492	AGTCACATAAAGTTGAGCTCATAGAGATAAATAGATTATACAAAGGAAGGCAACCAT	551
Db	727	AAAGAAACATGTTGAGTGATGGTGACTAAAGACATGGAAGATTAATAGGCCAAACCA	786
QY	552	CCATAGACCTGCACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACA	611
Db	787	GAACAGA--TGCTTGCACGCCATACCTTTTGATTCTTTTGTGATGTCAAAACAATACA	844
QY	612	AGTTGGATTTTGTGCTACAG---CTGAACACAGACGAGCAATCTCATGTATCCGATAC	668
Db	845	TAGAACATTTCTATTTCTACGTGTGCTGAGACAGGGCCACTCAATCTCATTTGACCCCAATAC	904
QY	669	ATGAATTCAAAGCCCTTCACAAATACACCAACACGACACAGAAGAGGATGTCAAGATGAAT	728
Db	905	ATGAATTCAAAGCCCTTTACAAACACAAAGAAAGCCACAGAAGAGGACATAAAATGAAT	964
QY	729	TTAGCAATGAGGTTTCCGATTGTGCTTCAGTTGTATGAATTCACGTACCAATGGGCACTA	788
Db	965	TTAGCAATGAGACTTCCGATTTTCGACGACTGTATGAATTCACGTACCAATGGGCACTA	1024
QY	789	TTCAATTTGGAGTCAAAAGACAAACCCCATGGGAAAATTTGTTGGCATCAAAAGTCACCAATG	848
Db	1025	TTCAATTTGGAGTAAAGACAAACCAACACGCGGAAAATTTGTTGTGTCAAAGTCACCAAGT	1083
QY	849	ATACCAAGGAAGCCCTCATTAACCATTTCAATCTGTATGATTAACAAGTATTTTGAAGACC	908
Db	1084	-----AAATATATCTTTGTTTAAACCATTTCAACACATGATCAGAAGTATTTTGAAGACA	1138
QY	909	ATCAAGTCCAAACGAAAGAGTGCATTCGAGAGCCAGATTTGTGGAAAGT-TTTACTG	967
Db	1139	GTGAATTCAGTGAAGCCAGGCTTTGATCCGGAAACACCGTTCTGTGAAGTCTTTTGTG	1198
QY	968	CCAAATAGTACTCTATCTGACAGATTTGTATTGGAATGGACATTAATTCACAGTCTCTCT	1027
Db	1199	CAGAAATAACACAATCTAATAGATTTGTCAATCAGGTAGATGTATTTCACGACATCTCT	1258
QY	1028	GAATGCCAATATGATTACTTCCAGATTAATAATGCAAAATACAAACAACAAATATGGGAA	1087
Db	1259	ATATGTCAAGAAAGTATTTCTCATTTATGATGCAAGTAGCAGCAGGCAAAACATGGAAA	1318
QY	1088	CAAAAGTAAAAAATTCCTATTTTGTGGAGATGGGACCGAGCTTAAGGACATTAACGAAA	1147
Db	1319	CAAAAGCAAGATACTTCAATCTTTGTAAAGAGAGAGCTAGCTCTAAGAAATATCCTGGGC	1378
QY	1148	AAT-----AAAGTTGATTTACAGACATTTAAAGCAGATTTTAAACACACTGGCA	1195
Db	1379	AATCCCAACCAACGGATAGAGATTTCAAGAAAATTTCTAGAGATTTAAGATGTGACA	1438
QY	1196	GAGTCCAGAAAAGCAGCAGAGAAAATAATTCAGAGCAAAAACAAATAAAAAGAAAGAGAG	1255
Db	1439	GGCTCTAGAAAAGCGCGCGAAGAGAACTTAGAATGGTAAAC---TAAGAAAGAGAGTCAA	1495
QY	1256	GGACCAAGTTGGTTAAATTAATTCACAGAGAAATCAAGATTTGTTAGATAATTTCACTAT	1315
Db	1496	GGACTTAAACTGTCTTAACTTTTGGACAGACACCGGCTCAGTTGATGAATCTTACTAT	1555
QY	1316	GAACAGTACATTTCTGTAAACAAATAAATTCACCCAGATCAAAACAAACACTTTAGATTTC	1375
Db	1556	GATTTGGTACATCTCTTTGTAAACAAATACCTGTGCACCAACTCAATTAGAGCACTTAGAGTTC	1615
QY	1376	CTGAAGGAATTAATGTTGTTGCTATATGGAGTTTGATCTGTAGTCTTAACATCAATGGA	1435
Db	1616	ATAAAGGAATGAATTTGTTGCTGTGTAGATTTTGATCTTATCTTCATCATCAGGGA	1675
QY	1436	GTGGTCAAAGCTTACAAAGAAAGCCGAGTAGCAAACTTCTACTTTTCCAAGTGTATGTAT	1495
Db	1676	GTGGTCAAAGCTTACAGAGAAAGCAGATAGCAAACTTCTACCTGCCAGTCACTATGAG	1735

QY 1496 GACAGAAACACACCAACCAATGAGACGATTCTTCTACTCTAAATCTTTTACCATCAA 1549

Db 1736 GAAAAAACTACCATAGCA--GAGAAATTTTACTCTGAACTTTTATGAGCAA 1786

RESULT 10

AL701060 762 bp mRNA linear EST 04-SEP-2003

LOCUS DXFZp686f114120.r1.686 (synonym: hlcc3) Homo sapiens cDNA clone

DEFINITION DXFZp686f114120.5', mRNA sequence.

ACCESSION AL701060

VERSION AL701060.1 GI:19621593

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 762).

AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and Wiemann,S.

TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.)

JOURNAL Unpublished (2001)

COMMENT Contact: MIPS

MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert.

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de, sequenced by MediGenomix (Martinried/Germany) within the cDNA sequencing consortium of the German Genome Project. No si sequence available.

This clone (DKFZp686f114120) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.

FEATURES

source

1..762

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp686f114120"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="686 (synonym: hlcc3)"

/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"

ORIGIN

Query Match 28.0%; Score 433.6; DB 9; Length 762;

Best Local Similarity 75.4%; Pred. No. 9.3e-68;

Matches 573; Conservative 0; Mismatches 169; Indels 18; Gaps 2

QY 697 CAATATACAGACACGCCACAGACGAGGTGTCAGATGCAATTTAGCAATGAGTTTCC 746

Db 1 CAACACAGAACACGCCACGGAAGTGGACATTAAGATGAAATTCAGCAATGAGTTTCC 60

QY 747 GATTTCCTTCAGCTTGATGAATTCACGTACCAATGGCACTATTCTATTTCGAGTCAAG 806

Db 61 GATTTCGATCAGCTTGTATGAATTCACGCCCAATGGCACCATCCATTTCGAGTCAAG 120

QY 807 ACAACCCCATGGGAATTTCTGGCATCAAGTCACCAATGATACCAAGAGAGCCCTCA 866

Db 121 ACAACCCCATGGGAATTTCTGGGTGTGGAATTCACCAAGT-----AGGCTGCTTCA 174

QY 867 TTAACCATTTCAATCTGATGATAACCAAGTATTTTGAAGACCATCAAGTCCCAACAGCAA 926

Db 175 TTGACCACTTCAATGATGATCAAAAAGTATTTTGAAGAAAGTGAGATCAATGAAACCA 234

QY 927 AGAAGTGCATTCGAGACCGCAGATTTGGGAAGTTTACTGCCAATAGTAGTCTCTATCTG 986

Db 235 AGAAGTGTATTGGGAGCCAGGTTTGTGGAAGTCCTCTGCAAGAACATACCATCTG 294

QY 987 ACAGATTTTGTATTGAAGTGACCATATTTCACAGTCTCTTGAATGCCAATGATGATTCT 104

```

|||||
295 ACAGATTGTCATTGAGTTGATATATTCACACACTATATGTAATGTAAGTATT 354
|||
1047 TCAGATTAAATGCAAAATTAACAACAAATATGGACAAAGTAAAAATCTTCAC 1106
|||
355 TCTACATTCAGATGCAAAATTTGTAAGATAAGATATGGAACAAACCAAAATCTTTAC 414
|||
1107 TATTTGTGGAGATGGACAGCTCTAAGACATTACGAAAAATA-----AAG 1154
|||
415 TCTTTGTAAGAGAGGGGTAGCTCTAGGATATCTGGCCATTCACAGCAACGGGATG 474
|||
1155 TTGATTTACAGACATTTAAAGCAGATTTTAAACACATGGCAGAGTCCGAAAGCAGAG 1214
|||
475 TAGATTTCAAGGCATTTTACAAAATTTAAAGTCACTGGTAGCATCTAGAAAGAGGCTG 534
|||
1215 AAGAAAAATTCAGAGCAAAACAAATATAAAAGAAAGAGAGGACCAAAAGTTGGTTAAAT 1274
|||
535 AAGAAGATATGGATGAAGCAATCAAGAAGAGAGTGAAGGATCAAGCTGGTTAAAC 594
|||
1275 TATTGACAGAAATCAAGATTTTGTAGATAATTCATATCAATGAAACAGTACATCTTTGAA 1334
|||
595 TTCTCATAGAAACCGAGACTCACTGGATAATTCATATCACTATGACTGGTACATCTTTGTA 654
|||
1335 CAAATAAATGCCACCCAGATCAAAACAACTTAGATTTCTGAGGAAATTAATAGT 1394
|||
655 CAAATAAATGCCATCCAAACCAATTAAGCATTTAGATTTTAAAGAAATTAATAGT 714
|||
1395 TTGCTGATTTGGAGTTTGTCTCTGAGTCTTAACATCAATGG 1434
|||
715 TTGCTGTTGGAGTTTGTCTCTGAGTCTTAATGATCAATGG 754

```

```

RESULT 11
BG615970 768 bp mRNA linear EST 18-APR-2001
LOCUS 602643421F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4774373 5',
DEFINITION mRNA sequence.
ACCESSION BG615970 GI:13667354
VERSION BG615970.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1647 row: j column: 06
High quality sequence stop: 671.

```

```

FEATURES
source
1..768
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4774373"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_61"
/notes="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgccg); Site_2: SfiI (ggccattagggc);
Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor

```

```

sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-DT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 clones
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match 26.8%; Score 415.4; DB 12; Length 768;
Best Local Similarity 92.8%; Pred. No. 1.7e-64;
Matches 491; Conservative 0; Mismatches 31; Indels 7; Gaps 5;

Qy 68 AGACTACTGGTATCAGATGCGCAAGCAACTTAACTTCCAGAAAATCAGATGATTGG 127
Db 240 AGTCAGTTTGGTATCAGATGCGCAAGCAACTTAACTTCCAGAAAATCAGATGATTGG 299
Qy 128 ACAAAAGAGATGTAATATCAGTGTGTAGAAAGTCATAAGATTCACCAAAAACACAGGAA 187
Db 300 ACAAAGAGATGTAATATCAGTGTGTAGAAAGTCATAAGATTCACCAAAAACACAGGAA 359
Qy 188 ATTTGACTGAACAAGACGTAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTT 247
Db 360 ATTTGACTGAACAAGACGTAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTT 419
Qy 248 GTTGATATGGCATCACACATGACACGAGTATTCAAATAGAGAACTATTTCAAGATTCG 307
Db 420 GTTGATATGGCATCACACATGACACGAGTATTCAAATAGAGAACTATTTCAAGATTCG 479
Qy 308 CGAAAAACAGCCATTGAAAGATTCGATTTCAGACATCTTAAGATGGGAAAGCCCGTAAAAAT 367
Db 480 CGAAAAACAGCCATTGAAAGATTCGATTTCAGACATCTTAAGATGGGAAAGCCCGTAAAAAT 539
Qy 368 GTCCTTAAAGACCAAACTGTCTCAAAAGAGAGTAGAGAACTTCAAGCAAAACAA 427
Db 540 GTCCTTAAAGACCAAACTGTCTCAAAAGAGAGTAGAGAACTTCAAGCAAAACAA 599
Qy 428 AAGGGT-AAAGAGAACCCAGATAT-GGCTAATCCGTC-CAATAGATGAGTAC--AACTGCTA 482
Db 600 AAGGGTCAAGAGAACCCAGATATGGCTAATCCGTCGCAATGAGTTCCTCACTGCTA 659
Qy 483 AAGTTTCTAGTCACTA--AAAGTTGAGTCTATAGAGATAAAATAGATTATACAGGA 540
Db 660 AAGTTTCTAGTCACTA--AAAGTTGAGTCTATAGAGATAAAATAGATTATACAGGA 719
Qy 541 AAGCAACCATCCATAGACCTGATGTATCATATCCATTTTGATGAA 589
Db 720 AAGCAACCATCCATAGACCTGATGTATCATATCCATTTTGATGAA 768

```

```

RESULT 12
CD706751 602 bp mRNA linear EST 25-JUN-2003
LOCUS EST23278 human nasopharynx Homo sapiens cDNA, mRNA sequence.
DEFINITION CD706751
ACCESSION CD706751
VERSION CD706751.1 GI:32237381
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and
Zeng, Y.-X.
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL Unpublished (2003)
COMMENT Contact: Yixin Zeng
Cancer Center
Sun Yat-sen University
651 Dongfeng Road East, Guangzhou 510060, China
Tel: 86-1380-9770-743
Fax: 86-20-8775-4506
Email: yxzeng@gzsums.edu.cn

```

```

FEATURES
  source
    Location/Qualifiers
      1..602
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /tissue_type="normal nasopharynx"
        /clone_lib="human nasopharynx"
        /notes="ESTs generated from a normal nasopharynx cdna library from southern Chinese"

ORIGIN
  Query Match      24.4%; Score 377.6; DB 14; Length 602;
  Best Local Similarity 97.7%; Pred. No. 9.9e-58;
  Matches 383; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 68 AGACTACTGGGTATCAGATGGCGAAGCAACTTAACTTCCAGAAAATACAGATGATGG 127
DB 211 AGTCAGTGGGTATCAGATGGCGAAGCAACTTAACTTCCAGAAAATACAGATGATGG 270
QY 128 ACAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAA 187
DB 271 ACAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAA 330
QY 188 ATTGTGCTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAGAACATCTT 247
DB 331 ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAGAACATCTT 390
QY 248 GTTGATATGGGCATCACACATGGACCACTATTCAAATAGAACTATTCAAAGAAATG 307
DB 391 GTTGATATGGGCATCACACATGGACCACTATTCAAATAGAACTATTCAAAGAAATG 450
QY 308 CGGAAACAGCCATTGAAGATTGCAATCAGACATCTTAAGATGGGAAGCCAGTAAAT 367
DB 451 CGGAAACAGCCATTGAAGATTGCAATCAGACATCTTAAGATGGGAAGCCAGTAAAT 510
QY 368 GCTCCTTAAGACCAAACTGTCTCAAGAGCAAGCTAGAGAACTTCAAAGCAAAACAA 427
DB 511 GCTCCTTAAGACCAAACTGTCTCAAGAGCAAGCTAGAGAACTTCAAAGCAAAACAA 570
QY 428 AAGGTTAAGAGAACCCAGATATGGCTAATCC 459
DB 571 AAGGTTAAGAGAACCCAGATATGGCTAATCC 602

RESULT 13
AW532890/c
LOCUS
DEFINITION
  UI-R-BUO-ang-d-03-0-UI.s1 UI-R-BUO Rattus norvegicus cdna clone
  UI-R-BUO-ang-d-03-0-UI 3', mRNA sequence.
ACCESSION
  AW532890
VERSION
  AW532890.1 GI:7175304
KEYWORDS
  EST.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
REFERENCE
  1 (bases 1 to 541)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  JOURNAL
  MEDLINE
  PUBMED
  9704477
  8889548
  Contact: Soares, MB
  Coordinated Laboratory for Computational Genomics
  University of Iowa
  375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
  Tel: 319 335 8250
  Fax: 319 335 9565
  Email: bento-soares@iowa.edu
  The sequence contained an oligo-dT track that was present in the
  oligonucleotide that was used to prime the synthesis of first

```

```

strand cdna and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cdna between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized Eye library cdna Library Preparation: M.B. Soares Lab
Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
  1..541
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /strain="Sprague-Dawley"
    /db_xref="taxon:10116"
    /clone="UI-R-BUO-ang-d-03-0-UI"
    /dev_stage="adult"
    /lab_host="DH10B (Life Technologies)"
    /clone_lib="UI-R-BUO"
    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BUO
    library is a subtracted library derived from a mixture of
    eye and ganglia tissues. For a detailed description of the
    library from which this clone was derived, please visit
    our web site at ratest.eng.uiowa.edu. The subtraction has
    been previously described in (Bonaldo, Lennon and Soares,
    Genome Research 6:791-806, 1996)
    TAG_TISSUE=Eye
    TAG_LIB=UI-R-BUO
    TAG_SEQ=CATTG"

```

ORIGIN

```

  Query Match      23.8%; Score 368.2; DB 10; Length 541;
  Best Local Similarity 80.0%; Pred. No. 4.8e-56;
  Matches 433; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 681 CTTTCAAAATACAGCAACAGCCAGAGAGGATGTCAAGATGAAATTTAGCAATGAGG 740
DB 541 CTTTAAACAAACACAGAGACACCCACAGAGAGGAGCATAAAGATGAAATTTAGCAACGAG 482
QY 741 TTTTCCGATTTGCTTCAGCTTGTATGAATTCAGTACCATGGCAGTACTTCAATTTGGAG 800
DB 481 TCTTCCGATTTGCTTTCAGCTTGTATGAATTCAGTACCATGGCAGTACTTCAATTTGGAG 422
QY 801 TCAAGAGCAAAACCCATGGGAAAATTTGTGGCATCAAAAGTCACCAATGATACCAAGGAG 860
DB 421 TCAAGAGCAAAACCCCAAGAGAGGATTTGTGGCATGGAGCTCACCCCGTCACTAAAGATG 362
QY 861 CCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCCAAAC 920
DB 361 CCTCATTTGACCACCTTCAATCTGATGATAGTCCAGTATTTTGGGTCCATCAAGTCCAAAC 302
QY 921 AAGCAAGAAGTGCATTTCGAGAGCCAGATTTGTGGAAGTTTACTGCCAAATAGTACTC 980
DB 301 AAGCAAGAAGTGCATTTCGAGAGCCAGATTTGTGGAAGTTTACTGCCAAATAGTACTC 242
QY 981 TATCTGACAGATTTGTTATTTGAAGTGGACATTTATCCAGAGTCTCTCGAATGCCAATATG 1040
DB 241 CATCTAATAAATTTGTTATTTGAGGTGGATATCATTTCCAATAATTTTCAAGATGCCAAGAG 182
QY 1041 ATTACTTCCAGATTTAAATGCAAAATTAACAACAATAATATGGGAACAAAAGTAAAAAT 1100
DB 181 ATTCTTCTCCAAATAAATTTGATTTGGCATAGGATGAACATGGCAACAATTTCCAAAT 122
QY 1101 TCTCACTATTTGTGCGAGATGGGACCAAGTCTTAAGGACATTTACGAAAAATTAAGTTGATT 1160
DB 121 ACTCGGTCTTTTGTGAGAGATGGTCCGAAAAGTAAGGACATCATAGGGAATTAAGCTGATT 62
QY 1161 TCAGAGCATTTAAAGCAGATTTTAAACACCTGGCAGAGCTCCAGAAACAGCAGAGAGAAA 1220
DB 61 TCAAGAGCTTTTAAAGTTAGACCTTAAAGGCATCTGGCAGACTCTTAGGAAAAAAGAAAAA 2
QY 1221 A 1221

```



```

Db      1 A 1

RESULT 14
AL700018      664 bp mRNA linear EST 04-SEP-2003
LOCUS      DKFZp686O10116 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION      DKFZp686O10116 5', mRNA sequence.
ACCESSION      AL700018
VERSION        AL700018.1 GI:19620551
KEYWORDS       EST.
SOURCE        Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 664)
AUTHORS        Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and
                Wiemann,S.
TITLE          EST (Duesterhoeft, et al.)
JOURNAL        Unpublished (1999)
COMMENT        Contact: MIPS
                MIPS
                Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
                This is the 5' sequence of the clone insert.
                Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
                consortium of the German Genome Project.
                No s1 sequence available.
                This clone (DKFZp686O10116) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source           1..664
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686O10116"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"

ORIGIN
Query Match      23.2%; Score 358.6; DB 9; Length 664;
Best Local Similarity 73.4%; Pred. No. 2.4e-54; Indels 18; Gaps 2;
Matches 492; Conservative 0; Mismatches 160;

QY      710 GAGGATGTCAAGATGAAATTTAGCAATGAGGTTTCCGATTGCTTCCAGTTGCTTCAGCTTGTATGAAT 769
Db      1 GTGGACATTAAGATGAAATTCAGCAATGAAGTCTCCGATTGTCATCAGCTTGTATGAAT 60
QY      770 TCACGTACCAATGGCAGCATTCATTTTGGAGTCAAGACCAACCCCATGGGAAATTTGTT 829
Db      61 TCACGCAACAAATGGCACCATTCCATTTTGGAGTCAAGGCAACCAACCATGGGAAATTTGTT 120
QY      830 GGCATCAAAAGTCACCAATGATACCAAGGAAGCCCTCATTAACCACTTTCAATCTCATGATA 889
Db      121 GGTGTGAATACACCAAGT-----AAGGCTGCCCTTCATTGACCACTTCATGTAATGATC 174
QY      890 AACAGTATTTGAAGACCATCAAGTCCACAGCAAGAAAGAGTGCATTCGAGAGCCCAAGA 949
Db      175 AAAAAAGTATTTGAAGAAAGTGAGATCAATGAAGCCAGAAAGTGTATTCGGAGCCCAAGG 234
QY      950 TTTTGGGAAGTTTACTGCCAAATAGTACTCTATCTGACAGATTGTTATTGAAGTGGAC 1009
Db      235 TTTTGGGAAGTCCCTCTCTCGAGAACATACACATCTGACAGATTGTCATTTGAAGTTGAT 294
QY      1010 ATTATTCACAGTTCTCTGAAATGCCAATATGATTACTTCCAGATTAAATGCAAAATTC 1069
Db      295 ACTATTCAAAACACTCTATATGTAATGATGAATATTTCTACATTCAGATGCAAAATTTGT 354

```

```

QY      1070 AACACAAAATATGGGAACAAAGTAAAAAATTTCTCACTATTGTTGGAGATGGGACCAGC 1129
Db      355 AAGATATAAATATGGAAACAAACCAAAATCTTTCACTGTTTGTAGAGAGGGGTAGC 414
QY      1130 TCTAAGGACATTAACGAAAATA-----AGTTGATTTCAGAGCATTTAAAGCA 1177
Db      415 TCTAGGGATATCTGTCGCCAATTCCAAGCAACGGGATGTAGATTTCAAGGCATTTTACAA 474
QY      1178 GATTTTAAAAACACTGGCAGAGTCCAGAAAAGCAGAGAGAAAATTCAGAGCAAAAACA 1237
Db      475 AATTTAAGTCACTGCTAGCATCTAGAAAAGAGGCTGAAGAAGAGTATGGAATGAAGCA 534
QY      1238 AATAAAAAAGAGAGAGGACCAAGTTGGTTAAATTTACAGAGAAATCAAGATTG 1297
Db      535 ATGAAGAAGAGAGTGAAGGACTAAAGCTGGTTAAACTTCTCATAGGAACCGAGACTCA 594
QY      1298 TTAGATAATTCATACATATCAACAGTACATTTCTGTAAACAAATAAATGCCACCCAGATCA 1357
Db      595 CTGGATAATTCATACATATCACTGCTGATCTTTGTAACAAATAAATGCCATCCANACCA 654
QY      1358 ACAAAACACT 1367
Db      655 AATAAGCACT 664

RESULT 15
BX473839      572 bp mRNA linear EST 04-SEP-2003
LOCUS      DKFZp686H08165 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION      DKFZp686H08165 5', mRNA sequence.
ACCESSION      BX473839
VERSION        BX473839.1 GI:31668083
KEYWORDS       EST.
SOURCE        Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 572)
AUTHORS        Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
                Fobo,G., Han,M. and Wiemann,S.
TITLE          EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
JOURNAL        Unpublished (2003)
COMMENT        Contact: MIPS
                MIPS
                Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
                This is the 5' sequence of the clone insert.
                Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
                sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
                consortium of the German Genome Project.
                No s1 sequence available.
                This clone (DKFZp686H08165) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source           1..572
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686H08165"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"

ORIGIN
Query Match      22.7%; Score 351.8; DB 13; Length 572;
Best Local Similarity 98.1%; Pred. No. 4.1e-53; Indels 0; Gaps 0;
Matches 356; Conservative 0; Mismatches 7;

QY      68 AGACTACTGGGATCAGATGGCAAGCAACTTAACCTTCAGAAATACAGATGTCG 127

```

Db	210	AGTCAGTTTGGTATCAGATGGCAAGCAACTTAACCTTCCAGAAAATACAGATGATTGG	269
Qy	128	ACAAAGAGGAGTGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACACACAGGAA	187
Db	270	ACAAAAGAGGATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAGAACACAGGAA	329
Qy	188	ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTT	247
Db	330	ATTTTGACTGAACAAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTT	389
Qy	248	GTTGATATGGGCATCACACATGGACCACTATTCAATAGAGAACTATTCAAGAAATTG	307
Db	390	GTTGATATGGGCATCACACATGGACCACTATTCAATAGAGAACTATTCAAGAAATTG	449
Qy	308	CGGAAAACAGCCATTGAAGATTGATTCAGACATCTTAAGATGGGAAAGCCCACTAAAAAT	367
Db	450	CGGAAAACAGCCATTGAAGATTGATTCAGACATCTTAAGATGGGAAAGCCCACTAAAAAT	509
Qy	368	GCTCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAACTTCAAAGCAAAACAA	427
Db	510	GCTCCTAAAGACCAAACTGTCTCAAAAGGACGTAGAGAACTTCAAAGCAAAACAA	569
Qy	428	AAG 430	
Db	570	AAG 572	

Search completed: April 26, 2004, 01:59:33
Job time : 2817.53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 76.8439 Seconds
(without alignments)
11186.557 Million cell updates/sec

Title: US-10-051-835-17

Perfect score: 1549

Sequence: 1 aaccacacagatatttcag.....ctctaaatctttaccatcaa 1549

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/ECTUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1462.4	94.4	4771	4	US-09-620-312D-121
2	68.6	4.4	7218	1	US-08-232-463-14
3	52.4	3.4	19124	2	US-08-487-826B-13
4	47.4	3.1	6326	4	US-10-204-708-58
5	47.4	3.0	1664976	4	US-08-916-421B-1
6	46.2	3.0	5562	4	US-10-204-708-63
7	45.6	2.9	832	4	US-09-621-976-2813
8	45.4	2.9	396	4	US-09-640-173-53
9	45.4	2.9	396	4	US-09-713-550-53
10	45.4	2.9	1248	2	US-08-897-340-5
11	45.4	2.9	1248	3	US-09-252-329-5
12	45.4	2.9	6243	2	US-09-056-078-1
13	44.8	2.9	832	4	US-09-621-976-2813
14	44.4	2.9	10467	4	US-10-204-708-2
15	44.2	2.9	6656	4	US-10-204-708-75
16	43.6	2.8	2447	2	US-09-014-969-14
17	43.6	2.8	5152	4	US-10-204-708-47
18	42.2	2.7	1806	4	US-09-601-198-27
19	42.2	2.7	1929	4	US-09-543-68A-3758
20	42.2	2.7	786431	4	US-09-751-389-3
21	41.8	2.7	6866	4	US-10-204-708-20
22	41.6	2.7	2223	1	US-08-257-073-4
23	41.2	2.7	11049	4	US-10-204-708-22
24	41.2	2.7	580073	4	US-08-545-528D-1
25	41.1	2.6	658	3	US-08-998-416-595
26	40.8	2.6	240	1	US-08-828-417-6
27	40.8	2.6	1405	1	US-07-710-361-3

28	40.8	2.6	2210	1	US-07-710-361-2
29	40.8	2.6	5562	4	US-10-204-708-64
30	40.6	2.6	5433	3	US-08-929-329-1
31	40.6	2.6	640681	4	US-09-790-988-1
32	40.4	2.6	1696	4	US-09-835-811-1
33	40.4	2.6	6669	4	US-10-204-708-6
34	40.4	2.6	11015	4	US-10-204-708-55
35	40.2	2.6	5852	1	US-07-867-108-2
36	40	2.6	4949	3	US-09-138-024-22
37	40	2.6	4949	4	US-09-404-066-22
38	40	2.6	4949	4	US-09-573-322-22
39	40	2.6	6070	4	US-10-204-708-10
40	39.8	2.6	2394	4	US-09-800-729-33
41	39.6	2.6	10619	4	US-10-204-708-4
42	39.4	2.5	8607	4	US-10-204-708-72
43	39.4	2.5	9347	4	US-10-204-708-36
44	39.2	2.5	2671	6	5168051-9
45	39.2	2.5	5181	1	US-08-257-073-10

ALIGNMENTS

RESULT 1
US-09-620-312D-121
; Sequence 121, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radjje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 121
; LENGTH: 4771
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3023)..(4771)
US-09-620-312D-121

Query Match 94.4%; Score 1462.4; DB 4; Length 4771;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 86 ATGCAAGCAACTTAACCTTCAGAAATACAGATGATTCGACAAAGAGGATGTAAT 145
DB 1 ATGCAAGCAACTTAACCTTCAGAAATACAGATGATTCGACAAAGAGGATGTAAT 60

Qy	146	CAGTGGTTGAAAGTCATATAGATTGACCAAAAACACAGGGAATTTTGTACTGACCAAGAC	205
Db	61	CAGTGGTTGAAAGTCATATAGATTGACCAAAAACACAGGGAATTTTGTACTGACCAAGAC	120
Qy	206	GTGAATGGAGCAGTCCTTGAAGTGGTTAAAAAAGAACATCTCTGTGATATGGGCATCA	265
Db	121	GTGAATGGAGCAGTCCTTGAAGTGGTTAAAAAAGAACATCTCTGTGATATGGGCATCA	180
Qy	266	CATGGACACGTTATCAAAATAGAAGAACTATTCAAGAAATTCGGGAANAACAGCCATTGAA	325
Db	181	CATGGACACGTTATCAAAATAGAAGAACTATTCAAGAAATTCGGGAANAACAGCCATTGAA	340
Qy	326	GATTTCGATTCCAGACATCTAAGATGGGAAAGCCCAAGTAAAAATGCTCCTAAAGACCAAACT	385
Db	241	GATTTCGATTCCAGACATCTAAGATGGGAAAGCCCAAGTAAAAATGCTCCTAAAGACCAAACT	300
Qy	386	GTGTCTCAAAAGGAACCTAGAGAACTTTCAAGACCAAAAACAAAAGGGTAAAGAGAACCCA	445
Db	301	GTGTCTCAAAAGGAACCTAGAGAACTTTCAAGACCAAAAACAAAAGGGTAAAGAGAACCCA	360
Qy	446	GATATGGCTAATCCGTCGCAATGAGTACAACCTGCTAAAGTTCTAAGTCACTATAAGTT	505
Db	361	GATATGGCTAATCCGTCGCAATGAGTACAACCTGCTAAAGTTCTAAGTCACTATAAGTT	420
Qy	506	GAGCTCATAGAAGTAATAATAGATTATACAAAAGGAGGCAACCTCCATAGACCTTGACA	565
Db	421	GAGCTCATAGAAGTAATAATAGATTATACAAAAGGAGGCAACCTCCATAGACCTTGACA	480
Qy	566	TGTTGATCATATCCATTGTAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT	625
Db	481	TGTTGATCATATCCATTGTAATTCAGTAATCCATATCGTTACAAGTTGGATTTTAGT	540
Qy	626	CTACAGCCTGAAAACAGACACAGGCAATCTCATTCATCCGATACATGAATTCAAAGCCTTC	685
Db	541	CTACAGCCTGAAAACAGACACAGGCAATCTCATTCATCCGATACATGAATTCAAAGCCTTC	600
Qy	686	ACAAATACAGCAACAGCCACAGAGAGGATGTCAGATGAAATTTAGCAATGAGGTTTTC	745
Db	601	ACAAATACAGCAACAGCCACAGAGAGGATGTCAGATGAAATTTAGCAATGAGGTTTTC	660
Qy	746	CGATTTCGCTTCAGCTTGTTATGAAATTCAGGTACCAATGGCATTTTCATTTTGGAGTCAAA	805
Db	661	CGATTTCGCTTCAGCTTGTTATGAAATTCAGGTACCAATGGCATTTTCATTTTGGAGTCAAA	720
Qy	806	GACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCCCAATGATACCAAGGAGGCCCTC	865
Db	721	GACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCCCAATGATACCAAGGAGGCCCTC	780
Qy	866	ATTAAACCATTTCAATCTCATGATATAAACCAAGTATTTGAAGACCAATCAAGTCCCAACAGCA	925
Db	781	ATTAAACCATTTCAATCTCATGATATAAACCAAGTATTTGAAGACCAATCAAGTCCCAACAGCA	840
Qy	926	AAGAAGTCGATTCGAGAGCCCAAGATTTGTGGAGGTTTATCTGCCAAATATGATCTCTATCT	985
Db	841	AAGAAGTCGATTCGAGAGCCCAAGATTTGTGGAGGTTTATCTGCCAAATATGATCTCTATCT	900
Qy	986	GACGATTTGTTATTGAAGTGGACATATTCCACAGTTCTCTGAATGCCAATATGATTAC	1045
Db	901	GACGATTTGTTATTGAAGTGGACATATTCCACAGTTCTCTGAATGCCAATATGATTAT	960
Qy	1046	TTCCAGATTAAATTCGAAAAATTAACAAACAAATATGGGAAACAAAGTAAATAAATTCCTCA	1105
Db	961	TTCCAGATTAAATTCGAAAAATTAACAAACAAATATGGGAAACAAAGTAAATAAATTCCTCA	1020
Qy	1106	CTATTGTGCGAGATGGGACACAGCTCTAAGGACATTACGAAAAATAAAGTTGATTTGAGT	1165
Db	1021	CTATTGTGCGAGATGGGACACAGCTCTAAGGACATTACGAAAAATAAAGTTGATTTGAGT	1080
Qy	1166	GCATTTAAAGCAGATTTTAAACACATCGGAGGTCCGAAAGCGAGAGAAATAATTC	1225
Db	1081	GCATTTAAAGCAGATTTTAAACACATCGGAGGTCCGAAAGCGAGAGAAATAATTC	1140
Qy	1226	AGAGCAAAAACAAATAAAAAAGAAAGAGAGGACCAAAAGTTGGTTAAATTTATTCACAGGA	1285

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US/07/935,313
; APPLICATION NUMBER:
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fis
US-08-232-463-14

```
Query Match      4.4%; Score 68.6; DB 1; Length 7218;
Best Local Similarity 3.6%; Pred. No. 2.7e-07;
Matches 14; Conservative 234; Mismatches 143; Indels 0; Gaps 0;

QY 46 ATATATGTAAGAAATTTGCAAGACTACTGGGTATCAGATGGCAAGCAACTTAACCT 105
DB 1457 AAGAGATAGAGAAATTTGTAACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1398

QY 106 TCCAGAAATACAGATGATGGACAAAGAGGATGTAATCAGTGTAGAAAAGTCATAA 165
DB 1397 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1338

QY 166 GATTGACAAACACAGAGGAATTTGACTGAACAAGAGTGATGAGCAGACTTGTAA 225
DB 1337 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1278

QY 226 GTGGTTAAAAAAGAACATCTTGTGATATGGGCATCACACATGAGCAGCATTCATAA 285
DB 1277 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1218

QY 286 AGAAGAACTATTCAAGAAATTCGGGAAACAGCCATTGAAGATTCCGATTCAGACATCAA 345
DB 1217 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1158

QY 346 GATGGGAAAGCCAGTAAATGCTCTTAAAGACCAAACTGTCTCAAAAGGAACGTAG 405
DB 1157 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1098

QY 406 AGAAACTTCAAGCAAAACAAAAGGGTAAA 436
DB 1097 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1067

RESULT 3
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: N1H121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match      3.4%; Score 52.4; DB 2; Length 19124;
Best Local Similarity 48.3%; Pred. No. 0.0036;
Matches 146; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 1062 AAAATTACAAACAACAAATATCGGAACCAAGTAAAGTAAAAAATTTCTCATTGTCGAGATG 1121
DB 15597 AAAAAATTAAGAAAAAATTTTATATAAAAAAATGATTATAAAAAAATAAAAAACAAA 15656

QY 1122 GCACGCTCTAAGCAGCATTCAGAAAAATTAAGTTGATTTCAGACCATTTAAAGCAGATT 1181
DB 15657 AGAAGAAAAAATAAACCATTTAAAAAATAAAAAAATATATATATATAAAAAAATAAAAAAAG 15716

QY 1182 TTAACACTGCGACAGTCCAGAAAGCAGACAGAGAAAAAATTCAGAGCAAAAAACAATA 1241
DB 15717 AAAAAATATATTAATAAATAAATAATATATATATATAAAAAAATAAAAAAATGTT 15776

QY 1242 AAAAAAGAAAGAGAGGACCAAAAGTTGGTTAAATTTTTCAGAGGAATTCAGATTGTTAG 1301
DB 15777 AAAAAAATAATATATACATAAAAAAATAAAAAAATTTTATTAATAAAAAAATAAAAAA 15836

QY 1302 ATAAATTCATCTATGACAGTACATCTTGTAAACAAATAAATGATGCCAGATCAAACAA 1361
DB 15837 AAAAAAATTAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 15896

QY 1362 AA 1363
DB 15897 AA 15898

RESULT 4
US-10-204-708-58/c
; Sequence 58, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 58
; LENGTH: 6326
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-58

Query Match      3.1%; Score 47.4; DB 4; Length 6326;
Best Local Similarity 44.9%; Pred. No. 0.04;
Matches 260; Conservative 0; Mismatches 316; Indels 3; Gaps 2;
```

```
QY 803 AAAGACACACCCATGGGAAATTTGTCATCAAGTCACCAATGATACCAA-GGAAGC 861
Db 2436 AAACAAAATCTAACTTTATTTTTCACACAAATTTCCCAATATCTCAATACGAAT 2377
QY 862 CTTCAATTAACCAATTTCAATCTGATGATGAATAACAAGTATTTTGAAGACCAATCAAGTCCAAACA 921
Db 2376 CTTTCTTTTCTTTTAAACGATACCTATATTTTAAACATCAATPACTACATTTTATATA 2317
QY 922 ASCAAAGAGTGCATTCGAGAGCCCAAGATTTGTGGAAGTTTACCTGCCAAATAGTACTCT 981
Db 2316 ACTAATAAAACAAATAAAATTTCTATATATTTTCTTTTACTTAACCTAACCAATATTAATA 2257
QY 982 ATCTGACAGATTTGTTATTTGAAGTGGACATTTTCCACAGTTCTCTGATGCCAATATGA 1041
Db 2256 TCATATATCAAAACACAAAAAATCTTTACAAAACCAACAATAATTCAAACACAAAAACCCA 2197
QY 1042 TTACTTCCAGATTAATAATCCAAATTTACAAACAAATATATGGGAACAAAGTAAATAAT 1101
Db 2196 TGTATCTCTTTTAAATAAATAATTAATCGCTAATTAACATATATTAATTAATA 2137
QY 1102 CTCACATTTTGTGCGAGATGGACCACTCTTAAGACATTTAGAAAAATAAAGTTGATTT 1161
Db 2136 ATAAAAACCTAACCCCAAAACTATATAACAAAAAACTAACTTACCTACCAATAATAT 2077
QY 1162 CAGAGCATTTAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAGAAA 1221
Db 2076 TAAAAACCTTAACCTAATAAACCTTAAACCT--CCAATCTATCTCCAAACCTAACGAAA 2019
QY 1222 ATTCAGAGCAAAACAAATTAATAAAGAAAGAGAGGAGCCAAAGTTGGTTAAATTAATTCAC 1281
Db 2018 CTTTCTCTCAAAACAAAAACGCAACACAAAACTTCAATATACGTATTTACTTCGATAA 1959
QY 1282 AGGAATCAAGATTTGTTAGATTAATTTCACTATGACAGTACATCTTGTAACCAATAA 1341
Db 1958 TAATATTTAATTTAATTTAATTAACCAAAACATAACCTCAAAACAAAAATAAATAAATAA 1899
QY 1342 ATGCCACCCAGATCAAAACAAACACTTAGATTTCCTGAA 1380
Db 1898 AATAAAATAAATAAATAAATAAATCTCTAATATCTACTTAA 1860

RESULT 5
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
```

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match 3.0%; Score 47; DB 4; Length 1664976;
Best Local Similarity 48.1%; Pred. No. 0.38;
Matches 164; Conservative 0; Mismatches 175; Indels 2; Gaps 1;

QY 1050 AGATTAAATGCARAATTACAAACAAACAAATATGGACAAAGTAAATAATTCCTCACTAT 1109
DB 844227 AAATTAAAGAGACTAATTCAAAAAATTAGAGAATTTAAAAAATATATACAA--CGTTTTTA 844170

QY 1110 TTGTCGAGATGGACAGCTCTAAGGACATTCGAAAAATTAAGTTGATTTTCAGAGCAT 1169
DB 844169 TTTTGTAGATGAGAAGTTTTTTAGACATCTTTTACAAAAAGATTTGGATGAATAATAA 844110

QY 1170 TTAAGCAGATTTTAAACACATGGCAGAGTCCAGAAAGCAGCAGAGAAAAATTCAGAG 1229
DB 844109 ATGAATATAAGGATTTTGACTTTATTTTACTACCTGGAGAGAAAAAGAAAAAC 844050

QY 1230 CAAAAACAATAAAAAAGAAAGAGAGGAGCCAAAGTTGGTTTAAATTTATGACAGAAATC 1289
DB 844049 CTAAGAAGTAAAAAAGAGAGATTAAAAAAGAAACCTGAAGAGAAAAATAGAAAAAGAAAA 843990

QY 1290 AAGATTGTTAGATAATTCATCTATGACAGTACATTTCTTGTACCAATTAATGCCACC 1349
DB 843989 TAGAATTTGTTAAGAAAGAGAAAAAGAACAAATTTATAAAAATCTGTGAAGATGTTG 843930

QY 1350 CAGATCAACAAAAACACTTAGATTTCTCTGAAGGAAATTTAAA 1390
DB 843929 AAGAGAAATTAACAACACTAATTTCCAAAGAGAAAAAA 843889
```

RESULT 6

```
US-10-204-708-63/c
; Sequence 63, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: FIEFENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63
```

```
Query Match 3.0%; Score 46.2; DB 4; Length 5562;
Best Local Similarity 49.2%; Pred. No. 0.075;
Matches 121; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 1145 AAAATAAAGTTGATTTTCAGAGCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGA 1204
DB 1625 AAAAAAATATAAAAAATAAAAATTAATAAATTTCTTAAACAAAAAATAAAAAA 1566

QY 1205 AAGACGACGAGAAAAAATTCAGACCAAAAAACAAATAAAAAAGAGAGGACCAAG 1264
DB 1565 AAACACTATAAAAAACAAAAAATAAATAAAAAAATAAATCTAACTAACAAAAATAA 1506
```

QY 1265 TTGGTTAAATTATTGACAGGAATCAAGATTGTTTGTAGATAATTCATCTATGACAGTAC 1324
Db 1505 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1446
QY 1325 ATTCTTGTACAAATAAATGCGACCCAGATCAAAACAAACATTTAGATTCTCTGAAGAA 1384
Db 1445 AATCAAAATAAATAAATCATCTCTAACCAGAAATAAATAAATAAATAAATAAATAA 1386
QY 1385 ATTAAA 1390
Db 1385 ATAAAA 1380

RESULT 7
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 2.9%; Score 45.6; DB 4; Length 832;
Best Local Similarity 15.3%; Pred. No. 0.053;
Matches 44; Conservative 132; Mismatches 111; Indels 1; Gaps 1;
QY 810 AACCCATCGGAAATTTGGATCAAAAGTCACCAATGATCCAGGAGCCCTCATTA 869
Db 87 MMWYWGWSYAMWTRTGTGYATYRMMYRWCNKKAYRTTCYSSKGTWTKR 146
QY 870 ACCATTTCATCTGATGATAAACAAGTATTTGAAGACCATCAAGTCCCAACAAAGA 929
Db 147 WKAATTTWWKTYAAATRYWMCWKWRASWYCWGWKARKWSTWRSYSAGAR 206
QY 930 AGTGATTCGAGCCAGCAAGATTGTGGAAGTTTACTGCGCAATAGTACTCTATCTGACA 989
Db 207 SAKRCCYSCSGAMSKWYKWWYRWRGNATGAGMKAWRASCMWRKRYAGKSKTSYKSMW 266
QY 990 GATTGTGTTAGTGGACATTATTCACAGTCTCTCAATGCCAATATGATTACTTCC 1049
Db 267 CWTBSWKYCYTKARWTGYCYRKGK-MWGRGHWYASKYMKWKRWWCWARMYRSTGTR 325
QY 1050 AGATTAATGCAAAATTCAACAAATAATGGGAACAAAGTAAAA 1097
Db 326 ASMWRRWRYTMMKWKYAWAARAARWMMWMAWRAAAATATAA 373

RESULT 8
US-09-640-173-53/c
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173

; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match 2.9%; Score 45.4; DB 4; Length 396;
Best Local Similarity 38.3%; Pred. No. 0.046;
Matches 142; Conservative
QY 923 GCAAGAGAGTGCAATTCGAGAGCCAGATTGTGGAAGTTTACTGCAATAGTACTCTA 982
Db 372 GCAAGNNNNNAANNAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 313
QY 983 TCTGACAGATTGTGTTATGAGTGGACATTATTCACAGTTCTCTGAATGCCAATATGAT 1042
Db 312 ANTATNAAGNANNATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 253
QY 1043 TACTTCAGATTAAATGCAAAATTACACAAATAATATGGGAACAAAGTAAAAATTC 1102
Db 252 TTAANAAGGNAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 193
QY 1103 TCATATTTCGCGAGATGGGACGACCTCTAAGGACATTACGAAATAATAAGTTGATT 1162
Db 192 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 133
QY 1163 AGACATTATAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAACGACGAGAGAAAA 1222
Db 132 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 73
QY 1223 TTCAGACAAAAACAATAAAAAAGAAAGAGAGGACCAAGTTGGTTAAATTTATTGACA 1282
Db 72 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 13
QY 1283 GGAATCAAGA 1293
Db 12 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA

RESULT 9
US-09-713-550-53/c
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-53

Query Match 2.9%; Score 45.4; DB 4; Length 396;
Best Local Similarity 38.3%; Pred. No. 0.046;
Matches 142; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 923 GCAAGAAGTGCATTCGAGAGCCGAAGATTGTTGGAGCTTTTACATGCGCAAAATAGTACTCTA 982
Db 372 GCAAGNNNNNAANNAGANN 313
QY 983 TCTGACAGATTGTTGTTGAGTGGACATTTATCCACAGTCTCTGAAATGCCAATATGAT 1042
Db 312 ANTATNAGNANNATNN 253
QY 1043 TACTTCCAGATTTAAATGCAAAATTCACACAAACAAATATGGGACAAAGTAAATAATTC 1102
Db 252 TTAATAAAGGNNAAAAAANNAANNTAAAAAANNAANNAANNAANNAANNAANNAANNA 193
QY 1103 TCATATTTGCGAGATGGGACAGCTCTAAGGACATTCAGRAAAATATAAGTTGATTC 1162
Db 192 AAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 133
QY 1163 AGAGCATTTAAAGCAGATTTTAAACACTGCGCAGAGTCCAGAAAGCAGCAGAGAAAA 1222
Db 132 AAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 73
QY 1223 TTCAGAGCAAAACAAATATAAAGAAAGAGAGAGAGGACCAAGTTGGTTTAAATTGACA 1282
Db 72 AAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 13
QY 1283 GGAATCAAGA 1293
Db 12 AAAAAAANNA 2

RESULT 10
US-08-897-340-5
; Sequence 5, Application US/08897340
; Patent No. 5955306
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,340
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/715,032
; FILING DATE: 17-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-005CP
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-897-340-5

Query Match 2.9%; Score 45.4; DB 2; Length 1248;
Best Local Similarity 51.9%; Pred. No. 0.069; 88; Indels 0; Gaps 0;
Matches 97; Conservative 2; Mismatches 88; Indels 0; Gaps 0;
QY 1059 TGCAGAAATTCACAAACAAATATGCGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1118
Db 1062 TGAAGATTTTTCAGTAAACATTTGAAGGCTGTAGACATTTTTCGATTTTGTACCTG 1121
QY 1119 ATGGGACGAGCTCTAAGGACATTCAGAAATATAAGTTGATTCAGAGCATTTAAAGCAG 1178
Db 1122 CAAGTGCCATCTTTAAGGGGGGAAATCATGAAGTCCACGTTACAGTAACTGTGTGTA 1181
QY 1179 ATTTTAAACACTGCGCAGAGTCCAGAAAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1238
Db 1182 TATTAATAAGTAAATTCAGTCMAAAAAAANNAANNAANNAANNAANNAANNAANNA 1241
QY 1239 ATAAAA 1245
Db 1242 AAAAAA 1248

RESULT 11
US-09-252-329-5
; Sequence 5, Application US/09252329
; Patent No. 6147192
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,329
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/897,340
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-005CP
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-252-329-5

Query Match 2.9%; Score 45.4; DB 3; Length 1248;
Best Local Similarity 51.9%; Pred. No. 0.069;
Matches 97; Conservative 2; Mismatches 88; Indels 0; Gaps 0;
QY 1059 TGCAGAAATTCACAAACAAATATGCGAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1118
Db 1062 TGAAGATTTTTCAGTAAACATTTGAAGGCTGTAGACATTTTTCGATTTTGTACCTG 1121

RESULT 14
US-10-204-708-2/c
; Sequence 2, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2003-05-06
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; LENGTH: 10467
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-2

Query Match 2.9%; Score 44.4; DB 4; Length 10467;
Best Local Similarity 48.6%; Pred. No. 0.26;
Matches 180; Conservative 0; Mismatches 186; Indels 4; Gaps 2;

QY	1035	AATATGATTCTTCAGATTAAATGCAAAATACCAACAAATATGGGACAAAGTA	1094
DB	8228	AAAAAAAAATCAATTCACAAAAAATTAATAAATTAATAAATTAATAAATTAATAA	8169
QY	1095	AAAAATTCATCTTCTGCGAGATGGGACGAGCTCTAAGGACATTCACGAAAAATAAG	1154
DB	8168	AAAAATTAACAAATAATATAATATATCCAAATATTCGTAATCGCAATAAAAAAT	8109
QY	1155	TTGATTTTCAGACATTTAAAGCAGATTTTAAACACTGGCAGAGTCAGAAAGCAGCAG	1214
DB	8108	ATAAACTCCAAACAAAAAATAATATCAAAACAA---ATATTTTAAATTTTCAAT	8052
QY	1215	AAGAAAAATTCAGACCAAAAAAATAAATAAAGAAAGAGAGGACCAAAAGTTGGTTAAAT	1274
DB	8051	AAATACTATTTACAAAAACACATTTTAAACATACATAAAAAAATAAATAAATAA	7992
QY	1275	TATTGACAGAAATCAAGTTTGTAGATAATTC-ATACTATGAACAGTACATCTTGTA	1333
DB	7991	TTTTTAAAAAATTCAAATATATTTTAAATCAAAAAAATAAATAAATAAATAAATA	7932
QY	1334	ACAAATAAATGCCACCAGATCAAAACAAACACTTTAGATTTCTCCTGAAGAAATTAATGG	1393
DB	7931	AACGATAAAAAAAACTTTAATATAAABACACTAAATTCCTCCAAAAAATAAATACAAAT	7872
QY	1394	TTTGCTGTAT 1403	
DB	7871	TAAAAATATAT 7862	

RESULT 15
US-10-204-708-75/c
; Sequence 75, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 75
; LENGTH: 6656
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-75

Query Match 2.9%; Score 44.2; DB 4; Length 6656;
Best Local Similarity 50.2%; Pred. No. 0.25;
Matches 135; Conservative 0; Mismatches 133; Indels 1; Gaps 1;

QY	1009	CATTATTCACAGTTCTCTGAATGCCAATATGATTACTTCCAGATTAATAATCGAAAAATTA	1068
DB	278	CGTCGTTACACTCCAACCTTAAACGACAAAAATAAAAAAACTATCTCAATAAATAATAA	219
QY	1069	CAACACAAAAATATGGACAAAGTAATAAATAATCTCACTATTTGTGCGAGATGGGACCCAG	1128
DB	218	ATAAATAAATAATACGAACTAATCAAAACTTATATACTATTTATATATATTTTAAAAATAAC	159
QY	1129	CTCTAAGGACATTACGAAAAATAAAGTTGATTTTCAGAGCAATTTAAAGCAGATTTTAAAC	1188
DB	158	CCTTAACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	99
QY	1189	ACTGGCAGATCCAGAAAGCAGAGAGAAAAATTCAGAGCAAAAAACAAATAAATAAACA	1248
DB	98	-CTATAAATAATTTATTAATAATTTATTACGAATAATAAACCAAAACAAATAAATAAATA	40
QY	1249	AAGAGAGGACCAAAAGTTGGTTAAATTTAT 1277	
DB	39	AAATTTTAACTTAATACTAATAAATAAT 11	

Search completed: April 26, 2004, 02:11:14
Job time : 81.8439 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:47:42 ; Search time 411.576 Seconds
(without alignments)
15988.422 Million cell updates/sec

Title: US-10-051-835-17
Perfect score: 1549
Sequence: 1 aacacatgatatttttcag.....ctctaaatcttaccatcaa 1549

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04.*

- 1: geneseqn1980s.*
- 2: geneseqn1990s.*
- 3: geneseqn2000s.*
- 4: geneseqn2001as.*
- 5: geneseqn2001bs.*
- 6: geneseqn2002as.*
- 7: geneseqn2003as.*
- 8: geneseqn2003bs.*
- 9: geneseqn2003cs.*
- 10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1549	100.0	1549	9 ADE25681	Ade25681 Human CDN
C 2	1470.4	94.9	4778	4 AAI24129	AAI24129 Probe #14
C 3	1470.4	94.9	4778	4 ABA69252	ABA69252 Human foe
C 4	1470.4	94.9	4778	4 AAI49425	AAI49425 Probe #18
C 5	1470.4	94.9	4778	4 ABA51244	ABA51244 Human bre
C 6	1470.4	94.9	4778	4 ABA36171	ABA36171 Probe #14
C 7	1470.4	94.9	4778	4 ABA33354	ABA33354 Human bon
C 8	1470.4	94.9	4778	4 AAK17545	AAK17545 Human bra
C 9	1470.4	94.9	4778	4 ABA42976	ABA42976 Human liv
C 10	1470.4	94.9	4778	5 AAI03703	AAI03703 Probe #96
C 11	1470.4	94.9	4778	6 ABA17437	ABA17437 Human gen
C 12	1462.4	94.4	4771	8 AAI58244	AAI58244 Human pol
C 13	1462.4	94.4	4771	8 ADB48211	ADB48211 Novel hum
C 14	1462.4	94.4	4771	8 ABA01147	ABA01147 Interfero
C 15	1043.4	67.4	1045	4 AAK77710	AAK77710 Human imm
C 16	672	43.4	705	6 ABA60169	ABA60169 Human can
C 17	636.2	41.1	6289	7 AAL60054	AAL60054 Human PCO
C 18	628.4	40.6	5870	4 AAI01146	AAI01146 Interfero
C 19	564.6	36.4	4185	4 AAI22155	AAI22155 Probe #12
C 20	564.6	36.4	4185	4 ABA67234	ABA67234 Human foe
C 21	564.6	36.4	4185	4 AAI47453	AAI47453 Probe #16
C 22	564.6	36.4	4185	4 ABA34329	ABA34329 Probe #12
C 23	564.6	36.4	4185	4 AAK41409	AAK41409 Human bon

C 24	564.6	36.4	4185	4 ABA41002	ABA41002 Human liv
C 25	564.6	36.4	4185	5 AAI07856	AAI07856 Probe #78
C 26	564.6	36.4	4185	6 ABA15415	ABA15415 Human gen
C 27	542.4	35.0	992	7 ABA63726	ABA63726 Human CDN
C 28	388	25.0	389	8 ACH31668	ACH31668 Human bon
C 29	361	23.3	644	4 AAS35870	AAS35870 Human car
C 30	361	23.3	644	4 AAK69748	AAK69748 Human imm
C 31	361	23.3	644	9 ADE45664	ADE45664 Human car
C 32	227.6	14.7	604	4 AAS27397	AAS27397 CDNA enco
C 33	227.6	14.7	604	9 ADB93575	ADB93575 Human CDN
C 34	138	8.9	356	4 AAS37122	AAS37122 Novel hum
C 35	73	4.7	251	4 AAS38441	AAS38441 Novel hum
C 36	71.6	4.6	12718	4 AAS35869	AAS35869 Human car
C 37	71.6	4.6	12718	4 AAK69747	AAK69747 Human imm
C 38	71.6	4.6	12718	9 ADE45663	ADE45663 Human car
C 39	61	3.9	14006	6 ABL33958	ABL33958 Human imm
C 40	58.6	3.8	6668	6 ABL33697	ABL33697 Human imm
C 41	58.4	3.8	2000	7 ADA71938	ADA71938 Rice gene
C 42	58.4	3.8	5750	4 AAS46708	AAS46708 Tumour su
C 43	58.4	3.8	5750	6 ABL34008	ABL34008 Human imm
C 44	56.8	3.7	6644	2 AAX33181	AAX33181 Base sequ
C 45	56.8	3.7	7372	2 AAX33182	AAX33182 Base sequ

ALIGNMENTS

RESULT 1
ID ADE25681 standard; CDNA; 1549 BP.
XX AC ADE25681;
XX DT 29-JAN-2004 (first entry)
XX DE Human CDNA differentially expressed in foam cells #85.
XX KW Human; ss; differential expression; foam cell; LPS; lipopolysaccharide;
XX KW cardiovascular disease; atherosclerosis.
XX OS Homo sapiens.
XX PN US2003194721-A1.
XX PD 16-OCT-2003.
XX PF 18-SEP-2002; 2002US-00247671.
XX PR 19-SEP-2001; 2001US-0323784P.
XX PA (INCY) INCYTE GENOMICS INC.
XX PI Mikita T, Shiffman D, Porter JG, Kaser MR;
XX DR WPI; 2003-875398/81.
XX PT Combination containing several polynucleotide that are differentially
XX PT expressed in foam cells and complements of the polynucleotides, useful
XX PT for diagnosing cardiovascular disease or atherosclerosis.
XX PS Claim 1; SEQ ID NO 85; 37pp; English.
XX CC The invention relates to a combination comprising several polynucleotides
XX CC having any one of 127 sequences (S1) such as the sequence of human
XX CC calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4
XX CC hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit
XX CC mRNA, etc., and their complements. The CDNAe are differentially expressed
XX CC in LPS (lipopolysaccharide)-treated foam cells. Also included are
XX CC obtaining an extended or full length gene from a library of nucleic acid
XX CC sequences, an expression vector containing the nucleic acids, a host cell
XX CC containing the vector, a purified polypeptide appearing as ADE25750 and
XX CC ADE25751, producing a protein by culturing the host cell, and a
XX CC composition comprising a purified antibody that specifically binds to the

CC proteins. The foam cell-expressed nucleic acids are useful for a high
CC throughput detection of differential expression of one or more
CC polynucleotides in a sample. The sample is from a subject with
CC atherosclerosis and comparison with a standard defines early, mid or late
CC stages of the disorder. The foam cell-expressed nucleic acids are useful
CC for high throughput screening of a library of molecules or compounds to
CC identify a ligand which binds a polynucleotide. The library is chosen
CC from DNA molecules, peptides, proteins and RNA molecules. The protein is
CC useful for a high throughput screening of library of molecules or
CC compounds to identify at least one ligand which specifically binds a
CC protein, for purifying a ligand from a sample for making an antibody. The
CC foam cell-expressed nucleic acids are useful for diagnosing
CC cardiovascular disorder. The foam cell-expressed nucleic acids are useful
CC as elements on a microarray which can be used for detecting related
CC polynucleotide in a sample, diagnosing cardiovascular disease,
CC atherosclerosis. The present sequence represents a cDNA whose expression
CC is upregulated in LPS treated foam cells.
XX
SQ Sequence 1549 BP; 609 A; 274 C; 287 G; 379 T; 0 U; 0 Other;

Query Match 100.0%; Score 1549; DB 9; Length 1549;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACACATCAGATATTTTCAGCACTAAAGAGATGTTTCCCCACATATATGTAAGAA 60
DB 1 AACACATCAGATATTTTCAGCACTAAAGAGATGTTTCCCCACATATATGTAAGAA 60
QY 61 ATTTCAGAGACTACTGGGTATCAGATGCGAAGCACTTAACCTCCAGAAATACAGA 120
DB 61 ATTTCAGAGACTACTGGGTATCAGATGCGAAGCACTTAACCTCCAGAAATACAGA 120
QY 121 TGATTGGCAAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAGATTGACCCAAACA 180
DB 121 TGATTGGCAAAAGAGGATGTAATCAGTGGTTAGAAAGTCATAGATTGACCCAAACA 180
QY 181 CAGGGAATTTTGACTGAACAGAGCTGATGAGACGCTCTGAGTGGTTAAAGAAAGA 240
DB 181 CAGGGAATTTTGACTGAACAGAGCTGATGAGACGCTCTGAGTGGTTAAAGAAAGA 240
QY 241 ACATCTTGTTGATATGGGCATCACACATGGACCGACCTATTCAAATAGAAAGCACTATTCAA 300
DB 241 ACATCTTGTTGATATGGGCATCACACATGGACCGACCTATTCAAATAGAAAGCACTATTCAA 300
QY 301 AGAATTGCGGAAACAGCCATTTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCGAG 360
DB 301 AGAATTGCGGAAACAGCCATTTGAAGATTCGATTCAGACATCTAAGATGGGAAGCCGAG 360
QY 361 TAAATATGCTCTTAAAGACCAAACTGTCTCAAAGGAACTAGAGAACTTCAAAGCA 420
DB 361 TAAATATGCTCTTAAAGACCAAACTGTCTCAAAGGAACTAGAGAACTTCAAAGCA 420
QY 421 AAAACAAAAGGGTAAAGAGAACCCAGATATGCTTAATCGCTGCAATGAGTACACTGC 480
DB 421 AAAACAAAAGGGTAAAGAGAACCCAGATATGCTTAATCGCTGCAATGAGTACACTGC 480
QY 481 TAAAGGTTCTAGTCTACTAAAAGTTGAGCTCATAGAAATAAATAGATTATACAAAGGA 540
DB 481 TAAAGGTTCTAGTCTACTAAAAGTTGAGCTCATAGAAATAAATAGATTATACAAAGGA 540
QY 541 AAGGCAACCATCCATAGACTGACATGTTGATATCATATCCATTTGATGAATTCAGTAATCC 600
DB 541 AAGGCAACCATCCATAGACTGACATGTTGATATCATATCCATTTGATGAATTCAGTAATCC 600
QY 601 ATATCGTTACAGTTGGATTTAGTCTACAGCTCGAAGCAGGCACTCTCATTTGA 660
DB 601 ATATCGTTACAGTTGGATTTAGTCTACAGCTCGAAGCAGGCACTCTCATTTGA 660
QY 661 TCCGATACATGATTAATCAAGCCCTTCAAAATACAGCAACAGCAAGAGGATGTCAA 720
DB 661 TCCGATACATGATTAATCAAGCCCTTCAAAATACAGCAACAGCAAGAGGATGTCAA 720
QY 721 GATGAATTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAA 780

DB 721 GATGAATTTAGCAATGAGGTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAA 780
QY 781 TGGCACTATTTTGGAGTCAAAAGCAAAACCCCATCGGAAATTTGTTGGCATCAAGT 840
DB 781 TGGCACTATTTTGGAGTCAAAAGCAAAACCCCATCGGAAATTTGTTGGCATCAAGT 840
QY 841 CACCAATGATACCAAGGAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTT 900
DB 841 CACCAATGATACCAAGGAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTT 900
QY 901 TGAAGACCATCAAGTCCAAACCAAGCAAAAGAGTGCATTCGAGAGCCAAAGATTTGCGAAGT 960
DB 901 TGAAGACCATCAAGTCCAAACCAAGCAAAAGAGTGCATTCGAGAGCCAAAGATTTGCGAAGT 960
QY 961 TTTACTGCCAAATAGTACTCTATCTGACAGATTTCTTATTTGAAGTGGACATTTATTCACA 1020
DB 961 TTTACTGCCAAATAGTACTCTATCTGACAGATTTCTTATTTGAAGTGGACATTTATTCACA 1020
QY 1021 GTTCTCTGAATGCCAATATGATTACTTCCAGATTTAAATGCAAAATTTACAAACAAAT 1080
DB 1021 GTTCTCTGAATGCCAATATGATTACTTCCAGATTTAAATGCAAAATTTACAAACAAAT 1080
QY 1081 ATGGGAACCAAGTAAAGAAATTTCTCACTATTTGCGAGATGGGACGCTCTTAAGACAT 1140
DB 1081 ATGGGAACCAAGTAAAGAAATTTCTCACTATTTGCGAGATGGGACGCTCTTAAGACAT 1140
QY 1141 TAGCAAAATTAAGTTGATTTCAGAGCATTTTAAAGCAGATTTTAAACACACTGGCAGAGTC 1200
DB 1141 TAGCAAAATTAAGTTGATTTCAGAGCATTTTAAAGCAGATTTTAAACACACTGGCAGAGTC 1200
QY 1201 CAGAAAGCAGCAGAGAAATTTTCAGCAAAACCAATTAAGAAAGAGAGGAGGACC 1260
DB 1201 CAGAAAGCAGCAGAGAAATTTTCAGCAAAACCAATTAAGAAAGAGAGGAGGACC 1260
QY 1261 AAAGTTGGTTAAATTTATGACAGGAAATCAAGATTTGTTAGATAATTCATCTATGAACA 1320
DB 1261 AAAGTTGGTTAAATTTATGACAGGAAATCAAGATTTGTTAGATAATTCATCTATGAACA 1320
QY 1321 GTAATTTCTTTGAACAAATTAATGCGACCCAGATCAACAAACACTTAGATTTCTGAA 1380
DB 1321 GTAATTTCTTTGAACAAATTAATGCGACCCAGATCAACAAACACTTAGATTTCTGAA 1380
QY 1381 GGAATTTAAATGTTTGTCTGTTTGGAGTTTGTATCTGAGTCTAAATCAATGAGTGGT 1440
DB 1381 GGAATTTAAATGTTTGTCTGTTTGGAGTTTGTATCTGAGTCTAAATCAATGAGTGGT 1440
QY 1441 CAAAGCTTCAAAAGAACCGGAGTAGCAAAACCTTCACTTTCCAAAGTGTATGTAGAAC 1500
DB 1441 CAAAGCTTCAAAAGAACCGGAGTAGCAAAACCTTCACTTTCCAAAGTGTATGTAGAAC 1500
QY 1501 GAAACACACACCAAAATGAGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
DB 1501 GAAACACACACCAAAATGAGACGATTTCTACTCTAAATCTTTTACCATCAA 1549

RESULT 2
AAI24129/c
ID AAI24129 standard; DNA; 4778 BP.
XX
AC AAI24129;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #14062 for gene expression analysis in human cervical cell sample.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX Homo sapiens.
XX OS
XX PN WO200157278-A2.
XX

PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00832366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX Claim 25; SEQ ID NO 14062; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human Hela cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
Query Match 94.9%; Score 1470.4; DB 4; Length 4778;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 78 GTATCAGATGGCAAGCAACTTACCTTCCAGAAATACAGATGATTGGCAAAAGAGG 137
DB 4778 GTATCAGATGGCAAGCAACTTACCTTCCAGAAATACAGATGATTGGCAAAAGAGG 4719
QY 138 ATGTAATCAGTGGTTAGAAAGTCTAAGTGGTTAAAGAAAGAAACATCTTCTGATATGG 197
DB 4718 ATGTAATCAGTGGTTAGAAAGTCTAAGTGGTTAAAGAAAGAAACATCTTCTGATGG 4659
QY 198 AACAAAGCGTGAATGGAGCAGCTTGAAGTGGTTAAAGAAAGAAACATCTTCTGATATGG 257
DB 4658 AACAAAGCGTGAATGGAGCAGCTTGAAGTGGTTAAAGAAAGAAACATCTTCTGATATGG 4599
QY 258 GCATCACACATGGACAGCTATTCAATAGAAAGAACTATTCAAGAAATGGCGAAACAG 317
DB 4598 GCATCACACATGGACAGCTATTCAATAGAAAGAACTATTCAAGAAATGGCGAAACAG 4539
QY 318 CCATTGAAGATTCGATTACAGACATCTAAGATGGGAAAGCCAGTAAAGATGCTCTAAAG 377
DB 4538 CCATTGAAGATTCGATTACAGACATCTAAGATGGGAAAGCCAGTAAAGATGCTCTAAAG 4479
QY 378 ACCAACTGTGTCTCAAAAGGAACTAGAGAACTTCAAGCAAAACAAAAGGTTAAAG 437
DB 4478 ACCAACTGTGTCTCAAAAGGAACTAGAGAACTTCAAGCAAAACAAAAGGTTAAAG 4419
QY 438 AGAACCCAGATATGCTATCGTCTGCAATGAGTACAACTGCTAAAGGTTCTAAGTCAC 497
DB 4418 AGAACCCAGATATGCTATCGTCTGCAATGAGTACAACTGCTAAGGTTCTAAGTCAC 4359
QY 498 TAAAGTGTGATCTATAGAAATAAATAGATTATACAAAGAAAGCAACATCCATAG 557
DB 4358 TAAAGTGTGATCTATAGAAATAAATAGATTATACAAAGAAAGCAACATCCATAG 4299
QY 558 ACCTGACATGTGTATCATATCCATTGTGATTCAGTATCCATATCCATGTTACAGTTGG 617

DB 4298 ACCTGACATGTGTATCATATCCATTGATGANTCAGTAAATCCATATCGTTACAGATTGG 4239
QY 618 ATTTTAGTCTACAGCTGAAACAGAGCCAGGCAATCTCATTGATCCGATACATGAATTC 677
DB 4238 ATTTTAGTCTACAGCTGAAACAGAGCCAGGCAATCTCATTGATCCGATACATGAATTC 4179
QY 678 AGCCCTTCACAAATACAGCAACAGCCACAGAGAGAGTGTCAAGATGAATTTAGCAATG 737
DB 4178 AAGCCCTTCACAAATACAGCAACAGCCACAGAGAGAGTGTCAAGATGAATTTAGCAATG 4119
QY 738 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTG 797
DB 4118 AGGTTTTCCGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTCATTTG 4059
QY 798 GAGTCAAGACACAAACCCCATGGAAATTTGTTGGCATCAAAAGTCAACCAATGATACCAAG 857
DB 4058 GAGTCAAGACACAAACCCCATGGAAATTTGTTGGCATCAAAAGTCAACCAATGATACCAAG 3999
QY 858 AAGCCCTTCATTAACCAATTTCAATCTGATGATAAAACAGTATTTTGAAGCAATCAAGTCC 917
DB 3998 AAGCCCTTCATTAACCAATTTCAATCTGATGATAAAACAGTATTTTGAAGCAATCAAGTCC 3939
QY 918 AACAGCAAGAGTGCATTCAGAGCCCAAGATTTGTGGAGTTTACTGCCCAATAGTA 977
DB 3938 AACAGCAAGAGTGCATTCAGAGCCCAAGATTTGTGGAGTTTACTGCCCAATAGTA 3879
QY 978 CTCTATCTCACAGATTTGTTTATTGAAGTGAACATTTATTCACAGTTCTCTGAATGCCAAT 1037
DB 3878 CTCTATCTCACAGATTTGTTTATTGAAGTGAACATTTATTCACAGTTCTCTGAATGCCAAT 3819
QY 1038 ATGATTACTTCAGATTAATAATGCAAAATACACAAATATATGGGACCAAGTAAAT 1097
DB 3818 ATGATTACTTCAGATTAATAATGCAAAATATACACAAATATATGGGACCAAGTAAAT 3759
QY 1098 AATTCTCACTATTGTCGAGATGGACCAAGTCTTAAGGACATTAACGAAAAATAAGTTG 1157
DB 3758 AATTCTCACTATTGTCGAGATGGACCAAGTCTTAAGGACATTAACGAAAAATAAGTTG 3699
QY 1158 ATTTGAGAGCATTTAAAGCAGATTTTAAACAACTGCGAGATCCAGAAAGCAGCAGAG 1217
DB 3698 ATTTGAGAGCATTTAAAGCAGATTTTAAACAACTGCGAGATCCAGAAAGCAGCAGAG 3639
QY 1218 AAAAAATTCAGAGCAAAAAACAAATAAAAAGAGAGGAGCCAAAGTTGGTTAAATAT 1277
DB 3638 AAAAAATTCAGAGCAAAAAACAAATAAAAAGAGAGGAGCCAAAGTTGGTTAAATAT 3579
QY 1278 TGACAGAAATCAAGATTTGTTAGATAATTCATATCTATGAAACAGTACATCTTTGTAACAA 1337
DB 3578 TGACAGAAATCAAGATTTGTTAGATAATTCATATCTATGAAACAGTACATCTTTGTAACAA 3519
QY 1338 ATAAATGCCACCCAGATCAACAAACACTTAGATTTCTGAGGAAATTTAAATGTTTG 1397
DB 3518 ATAAATGCCACCCAGATCAACAAACACTTAGATTTCTGAGGAAATTTAAATGTTTG 3459
QY 1398 CTGTATTGGAGTTTGATCTCTGAGTCTAACATCAATGGAGTGGTCAAGTTTCAAGAAA 1457
DB 3458 CTGTATTGGAGTTTGATCTCTGAGTCTAACATCAATGGAGTGGTCAAGTTTCAAGAAA 3399
QY 1458 GCCGAGTACCAACCTTCACTTCCAGGTATATGTAGAACAGAAACACACACCAAAATG 1517
DB 3398 GCCGAGTACCAACCTTCACTTCCAGGTATATGTAGAACAGAAACACACCAAAATG 3339
QY 1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
DB 3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307

RESULT 3
ABA69252/c
ID ABA69252 standard; DNA; 4778 BP.
XX
AC ABA69252;

3458 CTGTATTGGAGTTTGATCTGAGTCTAATCAATCGAGTGTCTCAAGAGCTTACAAAGAA 3399

1458 GCCGAGTAGCAACCTTCTCACTTCCAGTGTATATGTAGAACAGAAAACCAACCAATG 1517

3398 GCCGAGTAGCAACCTTCTCACTTCCAGTGTATATGTAGAACAGAAAACCAACCAATG 3339

1518 AGACGATTTCTACTCTAAATCTTTACCATCA 1549

3338 AGACGATTTCTACTCTAAATCTTTACCATCA 3307

RESULT 5

ABA51244/c

ID ABA51244 standard; DNA; 4778 BP.

XX AC ABA51244;

XX DT 01-FEB-2002 (first entry)

XX DE Human breast cell single exon nucleic acid probe #9939.

XX KW Human; microarray; single exon probe; gene expression; breast; disease;

XX KW cancer; ss.

XX OS Homo sapiens.

XX PN WO200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000662.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX PI WPI; 2001-496933/54.

XX DR New spatially-addressable set of single exon nucleic acid probes, useful

XX PT for measuring gene expression in sample derived from human breast,

XX PT comprises number of single exon nucleic acid probes.

XX PS Claim 4; SEQ ID NO 9939; 327pp + Sequence Listing; English.

XX CC The invention relates to a spatially-addressable set of single exon

XX CC nucleic acid probes for measuring gene expression in a sample derived

XX CC from human breast and Bt 474 cells. The method involves contacting the

XX CC probes with a collection of detectably labelled nucleic acids derived

XX CC from mRNA of human breast, and then measuring the label bound to each

XX CC probe of the microarray. The probes are useful for verifying the

XX CC expression of regions of genomic DNA predicted to encode proteins. They

XX CC are useful for gene discovery, and for determining predisposition and/or

XX CC prognosing breast disease. Gene expression analysis is useful for

XX CC assessing the toxicity of chemical agents on cells. The microarray of

XX CC this invention presents a far greater diversity of probes for measuring

XX CC gene expression, with far less bias than expressed sequence tag

XX CC microarrays. The method is suitable for rapid production of functional

XX CC information from genomic sequence. The present sequence is a single exon

XX CC nucleic acid probe of the invention. Note: The sequence data for this

XX CC patent did not form part of the printed specification, but was obtained

XX CC in electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;

Query Match	94.9%;	Score 1470.4;	DB 4;	Length 4778;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1471;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	78	GTATCAGAAATGGCAAGCAACTTAACTTCCAGAAAATACAGATGATTTGGCAAAAAGAGG	137	
DB	4778	GTATCAGAAATGGCAAGCAACTTAACTTCCAGAAAATACAGATGATTTGGCAAAAAGAGG	4719	
QY	138	ATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTTTGTGATATGG	197	
DB	4718	ATGTAAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTTTGTGATATGG	4659	
QY	198	AACAAGACGTGAATGGAGCAGCTCTTGAAGTGGTTAAAAAAGAAACATCTTGTTCGATATGG	257	
DB	4658	AACAAGACGTGAATGGAGCAGCTCTTGAAGTGGTTAAAAAAGAAACATCTTGTTCGATATGG	4599	
QY	258	GCATCACACATGGACACAGCTATTCAATAGAGAACTATTCAAGAAATTCGCGAAAAACAG	317	
DB	4598	GCATCACACATGGACACAGCTATTCAATAGAGAACTATTCAAGAAATTCGCGAAAAACAG	4539	
QY	318	CCATTGAAGATTTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAAATGCTCCTAAAG	377	
DB	4538	CCATTGAAGATTTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAAATGCTCCTAAAG	4479	
QY	378	ACCAAACTGTGCTCAAAAGGAACGTAGAGAACTTCAAGACAAAAACAAAGAGGTAAG	437	
DB	4478	ACCAAACTGTGCTCAAAAGGAACGTAGAGAACTTCAAGACAAAAACAAAGAGGTAAG	4419	
QY	438	AGAACCCAGATATGGCTAATCCGTCGCAATGAGTACAACCTGCTAAAGGTTCTAAGTCAC	497	
DB	4418	AGAACCCAGATATGGCTAATCCGTCGCAATGAGTACAACCTGCTAAAGGTTCTAAGTCAC	4359	
QY	498	TAAAGTTGAGCTCATAGAGATAAATAGATATATACAAAGGAAGCAACCATCCATAG	557	
DB	4358	TAAAGTTGAGCTCATAGAGATAAATAGATATATACAAAGGAAGCAACCATCCATAG	4299	
QY	558	ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGG	617	
DB	4298	ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGG	4239	
QY	618	ATTTTGTCTACAGCTGGAACAGGACCGCAATCTCATTGTCCGATACATGAATTCATCA	677	
DB	4238	ATTTTGTCTACAGCTGGAACAGGACCGCAATCTCATTGTCCGATACATGAATTCATCA	4179	
QY	678	AAGCCTTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG	737	
DB	4178	AAGCCTTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATG	4119	
QY	738	AGGTTTCCGATTTGCTTCAGCTTGTATGATATTCACGTACCAATGGCACATTCATTTTG	797	
DB	4118	AGGTTTCCGATTTGCTTCAGCTTGTATGATATTCACGTACCAATGGCACATTCATTTTG	4059	
QY	798	GAGTCAAAAGACAAAACCCCATCGGAAAAATTTGTTGGCATCAAAAGTCACCAATGATCAAGG	857	
DB	4058	GAGTCAAAAGACAAAACCCCATCGGAAAAATTTGTTGGCATCAAAAGTCACCAATGATCAAGG	3999	
QY	858	AAGCCCTCATTAACCATTTCAATCTGATGATATAACAGTATTTTGAAGACCATCAAGTCC	917	
DB	3998	AAGCCCTCATTAACCATTTCAATCTGATGATATAACAGTATTTTGAAGACCATCAAGTCC	3939	
QY	918	AACAAGCAAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAAAATAGTA	977	
DB	3938	AACAAGCAAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAAAATAGTA	3879	
QY	978	CTCTATCTGACAGATTTGTTTATTTGAAGTGGACATTTATTCACAGTTCTCTGAATGCCAAT	1037	
DB	3878	CTCTATCTGACAGATTTGTTTATTTGAAGTGGACATTTATTCACAGTTCTCTGAATGCCAAT	3819	
QY	1038	ATGATTACTTCCAGATTTAAAAATGCAAAATTTACAAACAAATAATATGGAAACAAAGTAAAA	1097	
DB	3818	ATGATTACTTCCAGATTTAAAAATGCAAAATTTACAAACAAATAATATGGAAACAAAGTAAAA	3759	
QY	1098	AATTCCTCACTATTTGTGCGAGATGGGACCACTCTTAAGGACATTTACGAAAAATAAAGTTG	1157	

Db 3758 AATTCACACTTGTGCGAGATGGGACAGCTCTAAGGACATACGMAAATAAAGTTG 3699
Qy 1158 ATTTGAGAGATTTAAAGCGATTTTAAACAACCTGGCAGAGTCAGAAAAGCGAGAG 1217
Db 3698 ATTTGAGAGATTTAAAGCGATTTTAAACAACCTGGCAGAGTCAGAAAAGCGAGAG 3639
Qy 1218 AAAAATTCAGAGCAAAACAAATAAAAGAGAGAGAGGACCAAAAGTTGGTTAAATAT 1277
Db 3638 AAAAATTCAGAGCAAAACAAATAAAAGAGAGAGAGGACCAAAAGTTGGTTAAATAT 3579
Qy 1278 TGACAGGAATCAAGATTTGTAGATTAATCAATCACTATGAACAGATCATTTCTTTGAACAA 1337
Db 3578 TGACAGGAATCAAGATTTGTAGATTAATCAATCACTATGAACAGATCATTTCTTTGAACAA 3519
Qy 1338 ATAATGCCCCAGATCAAAACAAACATTTAGATTTTCTTGAAGGAATTAATGTTG 1397
Db 3518 ATAATGCCACCCAGATCAAAACAAACATTTAGATTTTCTTGAAGGAATTAATGTTG 3459
Qy 1398 CTGTATTGGAGTTTGATCCTGAGTCTAAACATCAATGAGTGGTCAAAAGCTTCAAAAGAAA 1457
Db 3458 CTGTATTGGAGTTTGATCCTGAGTCTAAACATCAATGAGTGGTCAAAAGCTTCAAAAGAAA 3399
Qy 1458 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATGTAGAACAGAAACACACCAAAATG 1517
Db 3398 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATGTAGAACAGAAACACACCAAAATG 3339
Qy 1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
Db 3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307

RESULT 6
ABA36171/c
ID ABA36171 standard; DNA; 4778 BP.
XX ABA36171;
XX 23-JAN-2002 (first entry)
DE Probe #14637 for gene expression analysis in human heart cell sample.
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX Homo sapiens.
OS WO200157274-A2.
PN 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024283.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human hearts.
XX Claim 4; SEQ ID NO 14637; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
SQ
Query Match 94.9%; Score 1470.4; DB 4; Length 4778;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 78 GTATCAGATGGCAAGCAACTTAACCTTCCAGAAATACAGATGATGGACAAAAGAGG 137
Db 4778 GTATCAGATGGCAAGCAACTTAACCTTCCAGAAATACAGATGATGGACAAAAGAGG 4719
Qy 138 ATGTAAATCAGTGGTTAGAAAGTCAATAAGATTGACCAAAAACACAGGGAATTTTGAATG 197
Db 4718 ATGTAAATCAGTGGTTAGAAAGTCAATAAGATTGACCAAAAACACAGGGAATTTTGAATG 4659
Qy 198 AACAGAGCTGAATGAGCAGCTTTGAAGTGGTTAAAGAAAGACATCTTGTGTATATGG 257
Db 4658 AACAGAGCTGAATGAGCAGCTTTGAAGTGGTTAAAGAAAGACATCTTGTGTATATGG 4599
Qy 258 GCATCACACATGGACCAAGCTATTCAATAGAAAGAACTATTCAAGAAATGGGGAACAG 317
Db 4598 GCATCACACATGGACCAAGCTATTCAATAGAAAGAACTATTCAAGAAATGGGGAACAG 4539
Qy 318 CCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAG 377
Db 4538 CCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAG 4479
Qy 378 ACCAACTGTGTCTCAAAAGGAACTTAGAGAACTTTCAAGCAAAAACAAAAGGGTAAAG 437
Db 4478 ACCAACTGTGTCTCAAAAGGAACTTAGAGAACTTTCAAGCAAAAACAAAAGGGTAAAG 4419
Qy 438 AGAACCCAGATATGGCTAATCCGTCGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCAC 497
Db 4418 AGAACCCAGATATGGCTAATCCGTCGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCAC 4359
Qy 498 TAAAGTTGAGCTCATAGAAGATAAATAGATTATACAAAGAAAGCAACCATCCATAG 557
Db 4358 TAAAGTTGAGCTCATAGAAGATAAATAGATTATACAAAGAAAGCAACCATCCATAG 4299
Qy 558 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGG 617
Db 4298 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGG 4239
Qy 618 ATTTTAGTCTACAGCTGAACAGGACCCAGCAATCTCATTTGATCCGATACATGAATTC 677
Db 4238 ATTTTAGTCTACAGCTGAACAGGACCCAGCAATCTCATTTGATCCGATACATGAATTC 4179
Qy 678 AAGCCCTTCAAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAATTTAGCAATG 737
Db 4178 AAGCCCTTCAAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAATTTAGCAATG 4119
Qy 738 AGSTTTTCCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTG 797
Db 4118 AGSTTTTCCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTG 4059
Qy 798 GAGTCAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAAGTCCACCAATGATACCAAGG 857
Db 4058 GAGTCAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAAGTCCACCAATGATACCAAGG 3999
Qy 858 AAGCCCTCAATTAACCAATTTCAATCTGATGATGAATTTTGAAGCCCAAGTCC 917
Db 3998 AAGCCCTCAATTAACCAATTTCAATCTGATGATGAATTTTGAAGCCCAAGTCC 3939

QY 798 GAGTCAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCACCAATGATACCAAGG 857
Db 4058 GAGTCAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCACCAATGATACCAAGG 3999
QY 858 AAGCCCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTTTGAAGACCAATCAAGTCC 917
Db 3998 AAGCCCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTTTGAAGACCAATCAAGTCC 3939
QY 918 AACAGCAAAAGAGTGATTCGAGAGCCAGATTTGTGGAAGTTTACTGCCAAATAGTA 977
Db 3938 AACAGCAAAAGAGTGATTCGAGAGCCAGATTTGTGGAAGTTTACTGCCAAATAGTA 3879
QY 978 CTCATCTGCAGAGATTTGTTATTTGAAGTGCACATTTATTCACAGATTCCTCTGAATGCCAAT 1037
Db 3878 CTCATCTGCAGAGATTTGTTATTTGAAGTGCACATTTATTCACAGATTCCTCTGAATGCCAAT 3819
QY 1038 ATGATTACTTCCAGATTTAAATGCAAAATTAACAACAACAAATATGGAACAAAGTAAAA 1097
Db 3818 ATGATTACTTCCAGATTTAAATGCAAAATTAACAACAACAAATATGGAACAAAGTAAAA 3759
QY 1098 AATTCTCAGTATTTGCGAGATGGGACCGCTCTAAGGACATTTACGAAAAATAAAGTTG 1157
Db 3758 AATTCTCAGTATTTGCGAGATGGGACCGCTCTAAGGACATTTACGAAAAATAAAGTTG 3699
QY 1158 ATTTCCAGAGCATTTAAAGCAGATTTTAAACACTGCGAGATCCAGAAAAAGCAGAGAAG 1217
Db 3698 ATTTCCAGAGCATTTAAAGCAGATTTTAAACACTGCGAGATCCAGAAAAAGCAGAGAAG 3639
QY 1218 AAAAATTCAGAGCAAAACAAATTAACAAAGAGAGAGGACCAAACTGTTGTTAAATAT 1277
Db 3638 AAAAATTCAGAGCAAAACAAATTAACAAAGAGAGGACCAAACTGTTGTTAAATAT 3579
QY 1278 TGACAGGAAATCAAGATTTGTTAGATTAATTCATATGACAGTACATTTCTTTAAACAA 1337
Db 3578 TGACAGGAAATCAAGATTTGTTAGATTAATTCATATGACAGTACATTTCTTTAAACAA 3519
QY 1338 ATAAATGCCACCCAGATCAACAAACACATAGATTTTCTGAGGAAATTAATGTTG 1397
Db 3518 ATAAATGCCACCCAGATCAACAAACACATAGATTTTCTGAGGAAATTAATGTTG 3459
QY 1398 CTGTATTGAGTTTGATCTCTGAGTCTAACATCAATGGAGTGGTCAAAAGCTTACAAAGAAA 1457
Db 3458 CTGTATTGAGTTTGATCTCTGAGTCTAACATCAATGGAGTGGTCAAAAGCTTACAAAGAAA 3399
QY 1458 GCCAGTAGCAAAACCTTCACCTTCCAAAGTGTATATGTAGAACAGAAACACACCAAAATG 1517
Db 3398 GCCAGTAGCAAAACCTTCACCTTCCAAAGTGTATATGTAGAACAGAAACACACCAAAATG 3339
QY 1518 AGACGATTTCTACTTAATCTTTTACCATCAA 1549
Db 3338 AGACGATTTCTACTTAATCTTTTACCATCAA 3307

RESULT 8

AAK17545/c
ID AAK17545 standard; DNA; 4778 BP.
XX
AC AAK17545;
XX

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 17536.

XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.

XX Homo sapiens.

XX OS

XX WO200157275-A2.

XX XX

XX 09-AUG-2001.

PD

XX 30-JAN-2001; 2001WO-US000667.
XX PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
XX Example 4; SEQ ID NO 17536; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
XX Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
SQ

Query Match 94.9%; Score 1470.4; DB 4; Length 4778;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 78 GTATCAGATGCGCAAGCAACTTAACCTTCAGAAAAATACAGATGATTCGACAAAGAGG 137
Db 4778 GTATCAGATGCGCAAGCAACTTAACCTTCAGAAAAATACAGATGATTCGACAAAGAGG 4719
QY 138 ATGTAATCAGTGGTTAGAAAGTCTAAGATTGACCAAAAAACAGGGAATTTGACTG 197
Db 4718 ATGTAATCAGTGGTTAGAAAGTCTAAGATTGACCAAAAAACAGGGAATTTGACTG 4659
QY 198 AACAGAGCTGTAATGAGCAGTCTTGAAGTGGTTAAAAAAGAAACATCTTGTGTATATGG 257
Db 4658 AACAGAGCTGTAATGAGCAGTCTTGAAGTGGTTAAAAAAGAAACATCTTGTGTATATGG 4599
QY 258 GCATCACATCGACGACGCTATTCAAATAGAGAACTATTCAAAGATTCGCGAAAAACAG 317
Db 4598 GCATCACATCGACGACGCTATTCAAATAGAGAACTATTCAAAGATTCGCGAAAAACAG 4539
QY 318 CCATTGAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAG 377
Db 4538 CCATTGAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAG 4479
QY 378 ACCAACTGTGTCTCAAAAGGAACTAGAGAACTTCAAGCAAAAAACAAAGGGTTAAG 437
Db 4478 ACCAACTGTGTCTCAAAAGGAACTAGAGAACTTCAAGCAAAAAACAAAGGGTTAAG 4419
QY 438 AGAACCCAGATATGCTTAATCCGCTCGAATGAGTACAACTGCTAAAGGTTCTTAAGTCAAC 497
Db 4418 AGAACCCAGATATGCTTAATCCGCTCGAATGAGTACAACTGCTAAAGGTTCTTAAGTCAAC 4359
QY 498 TAAAGTTGAGTCTATAGAGATAAAATAGATTATACAAAGGAAAGGCAACCATCAATAG 557
Db 4358 TAAAGTTGAGTCTATAGAGATAAAATAGATTATACAAAGGAAAGGCAACCATCAATAG 4299
QY 558 ACCTGACATGTATCATATATCCATTTGATTCAGTAAATCCATATCGTTTACAAAGTTGG 617
Db 4298 ACCTGACATGTATCATATATCCATTTGATTCAGTAAATCCATATCGTTTACAAAGTTGG 4239
QY 618 ATTTTAGTCTACAGCTGAAACAGGACCAGGCAATCTCATTGATCCGATACATGAATTCAT 677

Db 4238 ATTTAGTCTACAGCTTGAACAGGACAGGCAATCTCATTGATCCGATACATGAAATCA 4179
 Qy 678 AAGCCTTCAAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAAATTTAGCAATG 737
 Db 4178 AAGCCTTCAAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAAATTTAGCAATG 4119
 Qy 738 AGTCTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGCACTATTCATTTTG 797
 Db 4118 AGTCTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGCACTATTCATTTTG 4059
 Qy 798 GAGTCAAAAGACAAACCCATCGGAAATTTGTTGGCATCAAAAGTCCCAATGATACCAAGG 857
 Db 4058 GAGTCAAAAGACAAACCCATCGGAAATTTGTTGGCATCAAAAGTCCCAATGATACCAAGG 3999
 Qy 858 AAGCCTTCAAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAAATTTAGCAATG 917
 Db 3998 AAGCCTTCAAAATACAGCAACAGCCACAGAGAGAGATGTCAAGATGAAATTTAGCAATG 3939
 Qy 918 AACCAAGCAAGAGAGTGCATTCGAGAGCCCAAGATTTGTTGGAAGTCTTACTGCCCAATAGTA 977
 Db 3938 AACCAAGCAAGAGAGTGCATTCGAGAGCCCAAGATTTGTTGGAAGTCTTACTGCCCAATAGTA 3879
 Qy 978 CTCTATCTACAGATTTGTTGATTTGAAGTGGACATTTATTCACAGTCTCTGATGCCAAT 1037
 Db 3878 CTCTATCTACAGATTTGTTGATTTGAAGTGGACATTTATTCACAGTCTCTGATGCCAAT 3819
 Qy 1038 ATGATTACTTCCAGATTTAAATGCAAAATTTCAACACAAATATGGGAACAAAGTAAAA 1097
 Db 3818 ATGATTACTTCCAGATTTAAATGCAAAATTTCAACACAAATATGGGAACAAAGTAAAA 3759
 Qy 1098 AATTCACATTTTGGGAGATGGACAGCTCTAAGACATTTAGCAAAATTAAGTTG 1157
 Db 3758 AATTCACATTTTGGGAGATGGACAGCTCTAAGACATTTAGCAAAATTAAGTTG 3699
 Qy 1158 AATTCAGAGCATTTAAAGCAGATTTTAAACACTGCGACAGTCCAGAAAGCAGCAGAG 1217
 Db 3698 AATTCAGAGCATTTAAAGCAGATTTTAAACACTGCGACAGTCCAGAAAGCAGCAGAG 3639
 Qy 1218 AAAAAATCAGAGCAAAACAAATATAAAGAGAGAGAGGAGCAAGTTGTTAAATAT 1277
 Db 3638 AAAAAATCAGAGCAAAACAAATATAAAGAGAGAGGAGGAGCAAGTTGTTAAATAT 3579
 Qy 1278 TGACAGGAATCAAGATTTGTTAGATTAATTCATCTATCAAGTACATCTCTGTTAAACAA 1337
 Db 3578 TGACAGGAATCAAGATTTGTTAGATTAATTCATCTATCAAGTACATCTCTGTTAAACAA 3519
 Qy 1338 ATAAATGCCACCCAGATCAAAACAAACATTTAGATTTCTGAGGAAATTAATGTTG 1397
 Db 3518 ATAAATGCCACCCAGATCAAAACAAACATTTAGATTTCTGAGGAAATTAATGTTG 3459
 Qy 1398 CTGTATTGGATTTGATCTCTGAGTCTAAACATCAATGAGTGGTCAAGCTTACAAAGAAA 1457
 Db 3458 CTGTATTGGATTTGATCTCTGAGTCTAAACATCAATGAGTGGTCAAGCTTACAAAGAAA 3399
 Qy 1458 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATCTAGAACAGAAAACCAACCAATG 1517
 Db 3398 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATCTAGAACAGAAAACCAACCAATG 3339
 Qy 1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
 Db 3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307

RESULT 9
 ABS42976/c
 ID ABS42976 standard; DNA; 4778 BP.
 XX ABS42976;
 AC ABS42976;
 XX 25-FEB-2003 (first entry)
 DT Human liver single exon probe, SEQ ID No 17966.
 XX

XX Human; single exon nucleic acid probe; liver; cirrhosis;
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
 XX coronary heart disease; ss.
 OS Homo sapiens.
 XX WO200157273-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000664.
 XX 04-FEB-2000; 2000US-0180312P.
 XX 26-MAY-2000; 2000US-0207456P.
 XX 30-JUN-2000; 2000US-00608408.
 XX 03-AUG-2000; 2000US-00632366.
 XX 21-SEP-2000; 2000US-0234687P.
 XX 27-SEP-2000; 2000US-0236359P.
 XX 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488898/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human adult liver.
 XX Claim 4; SEQ ID NO 17966; 658pp; English.
 XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
 XX measuring human gene expression in a sample derived from human adult
 XX liver, comprising one of 13109 defined nucleotide sequences given in the
 XX specification (or complements/ fragments). The probe hybridises at high
 XX stringency to a nucleic acid molecule expressed in the human adult liver.
 XX (I) may be used for predicting, measuring and displaying gene expression
 XX in samples derived from human adult liver. The genes identified may be
 XX involved in genetic liver diseases such as cirrhosis,
 XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
 XX associated with coronary heart disease. ABS25011-ABS51005 represent human
 XX liver single exon nucleic acid probes of the invention. Note: The
 XX sequence information for this patent does not appear in the printed
 XX specification but was obtained in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
 Qy Query Match 94.9%; Score 1470.4; DB 4; Length 4778;
 Db Best Local Similarity 99.9%; Pred. No. 0;
 Qy Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 78 GTATCAGATGGCAAGCAACTTAACTTCCAGAAATACAGATGATGGACAAAGAGG 137
 Qy 4778 GTATCAGATGGCAAGCAACTTAACTTCCAGAAATACAGATGATGGACAAAGAGG 4719
 Qy 138 ATGTAAATCAGTGGTTAGAAAAGTCATAAGATTCACAAACACACAGGAAATTTGACTG 197
 Db 4718 ATGTAAATCAGTGGTTAGAAAAGTCATAAGATTCACAAACACACAGGAAATTTGACTG 4659
 Qy 198 AACGAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAGAAAGACATCTTGTGATATGG 257
 Db 4658 AACGAGACGTGAATGGAGCAGTCTTGAAGTGGTTAAAGAAAGACATCTTGTGATATGG 4599
 Qy 258 GCATCACACATGGACCAAGCTTATCAATAGAAAGCTTATCAAGAAATTCGGAACACAG 317
 Db 4598 GCATCACACATGGACCAAGCTTATCAATAGAAAGCTTATCAAGAAATTCGGAACACAG 4539
 Qy 318 CCATTGAAGTTCGATTCAGACATCTAAGTGGGAAGCCCGATTAAGTCTCTCTAAAG 377
 Db 4538 CCATTGAAGTTCGATTCAGACATCTAAGTGGGAAGCCCGATTAAGTCTCTCTAAAG 4479

QY 378 ACCAACTGTCTCTCAAAAGGAAGCTAGAGAACTTCAAGCAAAAAACAAAGGGTAAAG 437
Db 4478 ACCAACTGTCTCTCAAAAGGAAGCTAGAGAACTTCAAGCAAAAAACAAAGGGTAAAG 4419
QY 438 AGAACCCAGATATGCTTAATCCGCTGCAATGAGTGAACAATGCTTAAAGTTCATCAAGTCAC 497
Db 4418 AGAACCCAGATATGCTTAATCCGCTGCAATGAGTGAACAATGCTTAAAGTTCATCAAGTCAC 4359
QY 498 TAAAGTTGAGCTCATAGAGATTAATAGATTAATACAAAGGAAGCAACCATCCATAG 557
Db 4358 TAAAGTTGAGCTCATAGAGATTAATAGATTAATACAAAGGAAGCAACCATCCATAG 4299
QY 558 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGTTTACAAGTTGG 617
Db 4298 ACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGTTTACAAGTTGG 4239
QY 618 ATTATTAGTCTACAGCTTAAACAGGACCAAGCAATCTCATTTGATCCGATCATCAATTCATCA 677
Db 4238 ATTATTAGTCTACAGCTTAAACAGGACCAAGCAATCTCATTTGATCCGATCATCAATTCATCA 4179
QY 678 AAGCCTTCACAAATACAGCAACAGCAGCAGAGAGGATGTCAAGATGAATTTAGCAATG 737
Db 4178 AAGCCTTCACAAATACAGCAACAGCAGCAGAGAGGATGTCAAGATGAATTTAGCAATG 4119
QY 738 AGTTTTCGATTTGCTTACGTTGTATGAATTCAGTACCAATGGCACTATTCATTTG 797
Db 4118 AGTTTTCGATTTGCTTACGTTGTATGAATTCAGTACCAATGGCACTATTCATTTG 4059
QY 798 GAGTCAAAAGCAAAACCCCATGGGAAATTTGTTGGCATCAAAAGTCACCAATGATACCAAGG 857
Db 4058 GAGTCAAAAGCAAAACCCCATGGGAAATTTGTTGGCATCAAAAGTCACCAATGATACCAAGG 3999
QY 858 AAGCCTCATTAACCAATTTCAATCTGATGATTAACAGATTTTGAAGACCATCAAGTCC 917
Db 3998 AAGCCTCATTAACCAATTTCAATCTGATGATTAACAGATTTTGAAGACCATCAAGTCC 3939
QY 918 AACAAAGCAAAAGTGCATTCAGAGCCAAAGATTTGTGGAAGTTTACTGCAATAGTA 977
Db 3938 AACAAAGCAAAAGTGCATTCAGAGCCAAAGATTTGTGGAAGTTTACTGCAATAGTA 3879
QY 978 CTCTATCTGACAGATTTGTTATGAGTGGACATTTATTCACAGTCTCTGAAATGCCAAT 1037
Db 3878 CTCTATCTGACAGATTTGTTATGAGTGGACATTTATTCACAGTCTCTGAAATGCCAAT 3819
QY 1038 ATGATTACTTCAGATTAATAAGTCAAAATTTACAAACAAATAATGGCAACAAAGTAAAA 1097
Db 3818 ATGATTACTTCAGATTAATAAGTCAAAATTTACAAACAAATAATGGCAACAAAGTAAAA 3759
QY 1098 AATTCTCACTATTTGTGAGATGGACAGCTCTAAGGACATTAAGAAAATAAAGTTG 1157
Db 3758 AATTCTCACTATTTGTGAGATGGACAGCTCTAAGGACATTAAGAAAATAAAGTTG 3699
QY 1158 ATTTCAGACATTTAAGCAGATTTTAAACACTGCGAGTCCAGAAAGCAGCAGAG 1217
Db 3698 ATTTCAGACATTTAAGCAGATTTTAAACACTGCGAGTCCAGAAAGCAGCAGAG 3639
QY 1218 AAAAAATTCAGAGCAAAACAAATAAAAGAAAGAGAGGAGCCAAAGTTGGTTAAATAT 1277
Db 3638 AAAAAATTCAGAGCAAAACAAATAAAAGAAAGAGAGGAGCCAAAGTTGGTTAAATAT 3579
QY 1278 TGAAGCAATCAAGATTTGTTAGATTAATTCATCTATCAAGCAGTACATCTCTGTAACAA 1337
Db 3578 TGAAGCAATCAAGATTTGTTAGATTAATTCATCTATCAAGCAGTACATCTCTGTAACAA 3519
QY 1338 ATAAATGCCACCCAGATCAAAACAAACATTTAGATTTCTCTGAAGGAAATTAATGGTTG 1397
Db 3518 ATAAATGCCACCCAGATCAAAACAAACATTTAGATTTCTCTGAAGGAAATTAATGGTTG 3459
QY 1398 CTGTATTGGAGTTGATCTCTGAGTCTTAACATCAATGAGTGTCAAGCTTCAAGAA 1457
Db 3458 CTGTATTGGAGTTGATCTCTGAGTCTTAACATCAATGAGTGTCAAGCTTCAAGAA 3399
QY 1458 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATAGAAACAGAAACCAACCAAAATG 1517

Db 3398 GCCGAGTAGCAAACTTCACTTTCCAAAGTGTATATGTAGAACAGAAACCAACCAAAATG 3339
QY 1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
Db 3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307
RESULT 10
AAI09703/c
ID AAI09703 standard; DNA; 4778 BP.
XX AAI09703;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #5694 used to measure gene expression in human breast sample.
XX KW Probe; human; breast disease; breast cancer; development disorder; ss;
XX KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WC200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US000661.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression in a human breast.
XX PS Claim 25; SEQ ID NO 9694; 322pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and prognosing diseases of the human breast,
XX CC particularly those diseases with polygenic aetiology. The diseases
XX CC include: breast cancer, disorders of development, inflammatory diseases
XX CC of the breast, fibrocytic changes, proliferative breast disease and non-
XX CC carcinoma tumours. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
Query Match 94.98; Score 1470.4; DB 5; Length 4778;
Best Local Similarity 99.94; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 78 GTATCAGATGGCAAGCAACTTAACTTCCAGAAAATACAGATGATGGCAAAAGAGG 137
Db 4778 GTATCAGATGGCAAGCAACTTAACTTCCAGAAAATACAGATGATGGCAAAAGAGG 4719
QY 138 ATGTAATCACTGTTGTAAGATCATAGATTGACCAAAACACAGGGAATTTTGACTG 197

Db 4718 ATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACACAGGAAATTTTGACTG 4659
Qy 198 AACAAAGCGTGAATGAGCAGTCTTCAAGTGGTTAAAAAAGAAACATCTTGTGTATGG 257
Db 4658 AACAGAGCGTGAATGAGCAGTCTTCAAGTGGTTAAAAAAGAAACATCTTGTGTATGG 4599
Qy 258 GCATCACAATGACGACGCTATTCAAAATAGAGAATCTTCAAGAAATTTGCGGAAAAACAG 317
Db 4598 GCATCACAATGACGACGCTATTCAAAATAGAGAATCTTCAAGAAATTTGCGGAAAAACAG 4539
Qy 318 CCATTGAAGATTTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAAAATGCTCTTAAG 377
Db 4538 CCATTGAAGATTTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAAAATGCTCTTAAG 4479
Qy 378 ACCAAATCTGTCTCAAAAGGACGTAGAGAACTTCAAGCAAAACAAAAGGGTAAAG 437
Db 4478 ACCAAATCTGTCTCAAAAGGACGTAGAGAACTTCAAGCAAAACAAAAGGGTAAAG 4419
Qy 438 AGAACCCAGATATGGCTTAATCCGCTGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCA 497
Db 4418 AGAACCCAGATATGGCTTAATCCGCTGCAATGAGTACAACTGCTAAAGGTTCTTAAGTCA 4359
Qy 498 TAAAGTTGAGCTCATAGAGATAAAATAGATTATCAAAAGGAAGCCAAACCATCATAG 557
Db 4358 TAAAGTTGAGCTCATAGAGATAAAATAGATTATCAAAAGGAAGCCAAACCATCATAG 4299
Qy 558 ACCTGACATGTTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGG 617
Db 4298 ACCTGACATGTTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTTGG 4239
Qy 618 ATTTTGTCTACAGCTGAAACAGGACCCAGGCAATCTCATTTGATCCGATACATGAATCA 677
Db 4238 ATTTTGTCTACAGCTGAAACAGGACCCAGGCAATCTCATTTGATCCGATACATGAATCA 4179
Qy 678 AAGCCTTCAAAATACAGCAACAGGACCAAGAGAGATGTCAGATGAATTTAGCAATG 737
Db 4178 AAGCCTTCAAAATACAGCAACAGGACCAAGAGAGATGTCAGATGAATTTAGCAATG 4119
Qy 738 AGSTTTTCCGATTGCTTCAAGTTGTATGAATTCAGTACCAATGCGCATTTCAATTTTG 797
Db 4118 AGSTTTTCCGATTGCTTCAAGTTGTATGAATTCAGTACCAATGCGCATTTCAATTTTG 4059
Qy 798 GAGTCAAGCAACACCCATGGGAAATTTGTCGCATCAAGTCACCAATGATACCAAG 857
Db 4058 GAGTCAAGCAACACCCATGGGAAATTTGTCGCATCAAGTCACCAATGATACCAAG 3999
Qy 858 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 917
Db 3998 AAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCC 3939
Qy 918 AACAGCAAAAGTGTATTCGAGGACCAAGATTTGTGGAATTTTACTGCCCAATAGTA 977
Db 3938 AACAGCAAAAGTGTATTCGAGGACCAAGATTTGTGGAATTTTACTGCCCAATAGTA 3879
Qy 978 CTCTATCTGACAGATTTGTTTATTTGAAGTGGACATTTTCCACAGTTCTCTGAATGCCAAT 1037
Db 3878 CTCTATCTGACAGATTTGTTTATTTGAAGTGGACATTTTCCACAGTTCTCTGAATGCCAAT 3819
Qy 1038 ATGATTAATCTCAGATTTAAATGCAAAATTAACAACAATAATGGAACCAAGTAATA 1097
Db 3818 ATGATTAATCTCAGATTTAAATGCAAAATTAACAACAATAATGGAACCAAGTAATA 3759
Qy 1098 AATTCTCACTATTGTCGAGATGGGACAGCTCTAAGGACATTAACAACAATAATGGAATG 1157
Db 3758 AATTCTCACTATTGTCGAGATGGGACAGCTCTAAGGACATTAACAACAATAATGGAATG 3699
Qy 1158 ATTTTCAGAGATTTAAAGCAGATTTTAAACATCTGCGAGAGTCCAGAAAGCAGCAGAG 1217
Db 3698 ATTTTCAGAGATTTAAAGCAGATTTTAAACATCTGCGAGAGTCCAGAAAGCAGCAGAG 3639
Qy 1218 AAAAATTCAGAGCAAAACAAATAAAAGAGAGAGGACCAAGTTGGTTAAATTAAT 1277
Db 3638 AAAAATTCAGAGCAAAACAAATAAAAGAGAGGAGGACCAAGTTGGTTAAATTAAT 3579

Qy 1278 TGACAGAAATCAAGATTTCTTAGATAAATTCATATATGAACAGTACATTTCTTTGAACAA 1337
Db 3578 TGACAGAAATCAAGATTTCTTAGATAAATTCATATATGAACAGTACATTTCTTTGAACAA 3519
Qy 1338 ATAAATGCCACCCAGATCAACAAACACATTTAGATTTCTTGAAGGAATTAATGTTTG 1397
Db 3518 ATAAATGCCACCCAGATCAACAAACACATTTAGATTTCTTGAAGGAATTAATGTTTG 3459
Qy 1398 CTGTATTGGAGTTTGTATCTCTGAGTCTAAATCAATCAATGAGTGGTCAAAAGTTTCAAAAGAAA 1457
Db 3458 CTGTATTGGAGTTTGTATCTCTGAGTCTAAATCAATCAATGAGTGGTCAAAAGTTTCAAAAGAAA 3399
Qy 1458 GCCAGTAGCAAAACCTTCACTTTCCAGTGTATATCTAGACAGAAACACACCCAAATG 1517
Db 3398 GCCAGTAGCAAAACCTTCACTTTCCAGTGTATATCTAGACAGAAACACACCCAAATG 3339
Qy 1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
Db 3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307

RESULT 11
ABS17437/c
ID ABS17437 standard; DNA; 4778 BP.
XX
AC ABS17437;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 17428.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW Chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenex syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO2000186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PI (MOLE-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PI WPI; 2002-114183/15.
XX
DR Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 4; SEQ ID NO 17428; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one
CC of 12614 nucleic acid sequences mentioned in the specification, or their

complements or the 12387 open reading frames derived from the 12614
probes. Also included are a microarray comprising the novel set of probes
; the novel set of probes which hybridise at high stringency to a nucleic
acid expressed in the human lung; measuring gene expression in a sample
derived from human lung, comprising (a) contacting the array with a
collection of detectably labeled nucleic acids derived from human lung
mRNA, and (b) measuring the label detectably bound to each probe of the
array; identifying exons in a eukaryotic genome, comprising (a)
algorithmically predicting at least one exon from genomic sequences of
the eukaryote; and (b) detecting specific hybridisation of detectably
labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
having a fragment identical to the predicted exon, the probe is included
in the above mentioned microarray; assigning exons to a single gene,
comprising (a) identifying exons from genomic sequence by the method
above and (b) measuring the expression of each of the exons in several
tissues and/or cell types using hybridisation to a single exon
microarrays having a probe with the exon, where a common pattern of
expression of the exons in the tissues and/or cell types comprising one
the exons should be assigned to a single gene; a peptide comprising one
of 12011 sequences, mentioned in the specification, or encoded by the
probes/open reading frames (ORF). The probes are used for gene expression
analysis, and for identifying exons in a gene, particularly using human
lung derived mRNA and for the study of lung diseases such as asthma, lung
cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
Kargener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
dyskinesia, pulmonary hypertension and hyaline membrane disease. The
present sequence is a single exon probe open reading frame of the
invention. Note: The sequence data for this patent did not form part of
the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 4778 BP; 1291 A; 911 C; 833 G; 1743 T; 0 U; 0 Other;
Query Match 94.9%; Score 1470.4; DB 6; Length 4778;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
78 GTATCAGATGCGAAGCACTTAACTTCCAGAAATACAGATGATGGACAAAGAGG 137
4778 GTATCAGATGCGAAGCACTTAACTTCCAGAAATACAGATGATGGACAAAGAGG 4719
138 ATGTAATCAGTGGTTAGAAAGTCATAAGATGACCAAAAAACACAGGGAAATTTGACTG 197
4718 ATGTAATCAGTGGTTAGAAAGTCATAAGATGACCAAAAAACACAGGGAAATTTGACTG 4659
198 AACAGACGTGAATGGAGCGCTTGAAGTGTATAAAAGAACATCTTTGATATGG 257
4658 AACAGACGTGAATGGAGCGCTTGAAGTGTATAAAAGAACATCTTTGATATGG 4599
258 GCATCACATGACGACGCTATTCAATAGAGAACTATTCAAGAAATTCGCGAAAAACAG 317
4598 GCATCACATGACGACGCTATTCAATAGAGAACTATTCAAGAAATTCGCGAAAAACAG 4539
318 CCATTGAAGATTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAATGCTCCTTAAG 377
4538 CCATTGAAGATTCGATTTCAGACATCTAAGATGGGAAGCCAGTAAATGCTCCTTAAG 4479
378 ACCAACTGTGCTCAAAAGGACGTAGAGAACTTCAAGCAAAAAACAAAGGGGTAAAG 437
4478 ACCAACTGTGCTCAAAAGGACGTAGAGAACTTCAAGCAAAAAACAAAGGGGTAAAG 4419
438 AGAACCCAGATATGGCTAATCCGCTGCAATGAGTCAAACTGCTAAAGGTTCTTAAGTCAC 497
4418 AGAACCCAGATATGGCTAATCCGCTGCAATGAGTCAAACTGCTAAAGGTTCTTAAGTCAC 4359
498 TAAAGTTGAGCTCATAGAGATATAATAGATTATACAAAGGAAGCAACCATCCATAG 557
4358 TAAAGTTGAGCTCATAGAGATATAATAGATTATACAAAGGAAGCAACCATCCATAG 4299

558 ACCTGACATGTGTATCATATCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGG 617
4298 ACCTGACATGTGTATCATATCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGG 4239
618 ATTTTGTAGTCTACAGCCTGAAAACAGGACCAAGCAATCTCATTTGATCGGATACATGAATTC 677
4238 ATTTTGTAGTCTACAGCCTGAAAACAGGACCAAGCAATCTCATTTGATCGGATACATGAATTC 4179
678 AAGCTTTCACAAATACAGCAACAGCCACAGAGAGATGTCAAGATGAATTTAGCAATG 737
4178 AAGCTTTCACAAATACAGCAACAGCCACAGAGAGATGTCAAGATGAATTTAGCAATG 4119
738 AGGTTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTG 797
4118 AGGTTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTG 4059
798 GAGTCAAGACAAACCCCATGGGAAAATTTGTCGATCAAGTCACCAATGATACCAAG 857
4058 GAGTCAAGACAAACCCCATGGGAAAATTTGTCGATCAAGTCACCAATGATACCAAG 3999
858 AAGCCCTCATTAACCACTTCAATCTCATGATAACCAAGTATTTTGAAGACCATCAAGTCC 917
3998 AAGCCCTCATTAACCACTTCAATCTCATGATAACCAAGTATTTTGAAGACCATCAAGTCC 3939
918 AACAGCAAAAGAGTGCATTCGAGAGCCAAAGATTTGTGGAAGTTTACTGCCAAAATAGTA 977
3938 AACAGCAAAAGAGTGCATTCGAGAGCCAAAGATTTGTGGAAGTTTACTGCCAAAATAGTA 3879
978 CTCTCTCTGACAGATTTGTTTATTTGAAGTGACATTTCCACAGTCTCTCTGAATGCCAAT 1037
3878 CTCTCTCTGACAGATTTGTTTATTTGAAGTGACATTTTCCACAGTCTCTCTGAATGCCAAT 3819
1038 ATGATTAATCTCCAGATTAATAATCAAAATTTACCAACCAAAATATGGGAACAAAGTAAAA 1097
3818 ATGATTAATCTCCAGATTAATAATCAAAATTTACCAACCAAAATATGGGAACAAAGTAAAA 3759
1098 AATTTCTCACTATTTGTGCGAGATGGGACCAAGCTTAAGGACATTTAGAAAAATAAAGTTG 1157
3758 AATTTCTCACTATTTGTGCGAGATGGGACCAAGCTTAAGGACATTTAGAAAAATAAAGTTG 3699
1158 ATTTTCAGAGCATTTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAAGCAGAGAAG 1217
3698 ATTTTCAGAGCATTTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAAGCAGAGAAG 3639
1218 AAAAAATTCAGAGCAAAAAACAAATTAAGAAAGAGAGGAGGACCAAGTTGGTTAAATAT 1277
3638 AAAAAATTCAGAGCAAAAAACAAATTAAGAAAGAGAGGAGGACCAAGTTGGTTAAATAT 3579
1278 TGACAGAAATCAAGATTTGTAGATAATTCATACATATGAACAGTACATTTCTTTGTAACAA 1337
3578 TGACAGAAATCAAGATTTGTAGATAATTCATACATATGAACAGTACATTTCTTTGTAACAA 3519
1338 ATAAATGCCACCCAGATCAAAACAAACATAGATTTCTCTGAAGGAATTAATGGTTTG 1397
3518 ATAAATGCCACCCAGATCAAAACAAACATAGATTTCTCTGAAGGAATTAATGGTTTG 3459
1398 CTGTATTTGGAGTTTGTGCTGAGTCTAAACATCAATGGAGTGGTCAAGAGTTTAAATGGTTG 1457
3458 CTGTATTTGGAGTTTGTGCTGAGTCTAAACATCAATGGAGTGGTCAAGAGTTTAAATGGTTG 3399
1458 GCCGAGTAGCAAAACCTTCACTTTCCAAAGTGTATATGATAGAACAGAAACCAACCAATG 1517
3398 GCCGAGTAGCAAAACCTTCACTTTCCAAAGTGTATATGATAGAACAGAAACCAACCAATG 3339
1518 AGACGATTTCTACTCTAAATCTTTTACCATCAA 1549
3338 AGACGATTTCTACTCTAAATCTTTTACCATCAA 3307
RESULT 12
RAI58244
ID AAI58244 standard; cdna; 4771 BP.
XX

AA158244;
 22-OCT-2001 (first entry)
 Human polynucleotide SEQ ID NO 447.
 Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 peripheral nervous system; neuropathy; central nervous system; CNS;
 Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 leukaemia; ss.
 OS Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US034263.
 XX 23-DEC-1999; 99US-00471.275.
 XX 21-JAN-2000; 2000US-00488725.
 XX 25-APR-2000; 2000US-0052317.
 XX 20-JUN-2000; 2000US-00598042.
 XX 19-JUL-2000; 2000US-00620312.
 XX 03-AUG-2000; 2000US-00653450.
 XX 14-SEP-2000; 2000US-00662191.
 XX 19-OCT-2000; 2000US-00693036.
 XX 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AM39088.
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 PT Claim 1; SEQ ID NO 447; 10078pp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AM38842-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX Sequence 4771 BP; 1740 A; 833 C; 909 G; 1289 T; 0 U; 0 Other;
 SQ

Query Match 94.4%; Score 1462.4; DB 4; Length 4771;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 86 ATGGCAAGCACTTAACCTTCAGAAATACAGATGATGACAAAGAGAGATTAAT 145
 DB 1 ATGGCAAGCACTTAACCTTCAGAAATACAGATGATGACAAAGAGAGATTAAT 60
 QY 146 CAGTGGTTAGAAAGTCATAGATTTGACCAAAACACAGGGAAATTTTGACTGAACAAGAC 205

61 CAGTGGTTAGAAAGTCATAGATTTGACCAAAACACAGGGAAATTTTGACTGAACAAGAC 120
 QY
 206 GTGAATGGAGCAGTCTTGAAGTGGTTTAAAAAAGAACATCTTGTGTGATATGGGCATCACA 265
 DB
 121 GTGAATGGAGCAGTCTTGAAGTGGTTTAAAAAAGAACATCTTGTGTGATATGGGCATCACA 180
 QY
 266 CATGGACCAAGCTATTCAAATAGAAAGAACTATTCAAAGAAATTCGGGAAAAACAGCCATTGAA 325
 DB
 181 CATGGACCAAGCTATTCAAATAGAAAGAACTATTCAAAGAAATTCGGGAAAAACAGCCATTGAA 240
 QY
 326 GATTCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAAAATGCTCTTAAAGACCAAACT 385
 DB
 241 GATTCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAAAATGCTCTTAAAGACCAAACT 300
 QY
 386 GTGTCTCAAAAGGAACCTAGAGAACTCTCAAAAGCAAAAACAAAGGTTAAAGAGAACCCA 445
 DB
 301 GTGTCTCAAAAGGAACCTAGAGAACTCTCAAAAGCAAAAACAAAGGTTAAAGAGAACCCA 360
 QY
 446 GATATGGCTAATCCGTCTGCAATGAGTACAACCTGCTAAAGGTTCTAAGTCACTAAAGTT 505
 DB
 361 GATATGGCTAATCCGTCTGCAATGAGTACAACCTGCTAAAGGTTCTAAGTCACTAAAGTT 420
 QY
 506 GAGCTCATAGAGATAAATAGATATATACAAAGGAAGCAACCACTCCATAGACCTTGACA 565
 DB
 421 GAGCTCATAGAGATAAATAGATATATACAAAGGAAGCAACCACTCCATAGACCTTGACA 480
 QY
 566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTGTGATTTAGT 625
 DB
 481 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTGTGATTTAGT 540
 QY
 626 CTACAGCCTGAAACAGGACCCGCAATCTCATTGTATCCGATCCGATCAATGAATTCAAAGCCTTC 685
 DB
 541 CTACAGCCTGAAACAGGACCCGCAATCTCATTGTATCCGATCCGATCAATGAATTCAAAGCCTTC 600
 QY
 686 ACAATATACGCAACACGCCACAGAAAGAGATGTCAAGATGAAATTTAGCAATAGGTTTTC 745
 DB
 601 ACAATATACGCAACACGCCACAGAAAGAGATGTCAAGATGAAATTTAGCAATAGGTTTTC 660
 QY
 746 CGATTTGCTTCAGCTTGTATGATTTCACTACCAATGCGCACTATTCATTTCGGAGTCAAA 805
 DB
 661 CGATTTGCTTCAGCTTGTATGATTTCACTACCAATGCGCACTATTCATTTCGGAGTCAAA 720
 QY
 806 GACAAACCCCATGGGAAAATTTGTTGGCATCAAAGTCACCAATGATACCAAGAGAGCCCTC 865
 DB
 721 GACAAACCCCATGGGAAAATTTGTTGGCATCAAAGTCACCAATGATACCAAGAGAGCCCTC 780
 QY
 866 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCCAACAGCA 925
 DB
 781 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCATCAAGTCCCAACAGCA 840
 QY
 926 AAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAATAGTACTCTATCT 985
 DB
 841 AAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAATAGTACTCTATCT 900
 QY
 986 GACAGATTTGTTTGAAGTGGACATTTCCACAGTCTCTCTGAATGCCAATAGTTAT 1045
 DB
 901 GACAGATTTGTTTGAAGTGGACATTTCCACAGTCTCTCTGAATGCCAATAGTTAT 960
 QY
 1046 TTCAGATTTAAAATGCAAAATTTACAAACAATAATGGGAAACAAAGTAAAAAATTTCTCA 1105
 DB
 961 TTCAGATTTAAAATGCAAAATTTACAAACAATAATGGGAAACAAAGTAAAAAATTTCTCA 1020
 QY
 1106 CTATTTGTGGAGATGGGACAGCTCTAAGGACATTTACGAAAAATAAAGTTGATTTTCA 1165
 DB
 1021 CTATTTGTGGAGATGGGACAGCTCTAAGGACATTTACGAAAAATAAAGTTGATTTTCA 1080
 QY
 1166 GCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAAGCAGACAGAAAAATTC 1225
 DB
 1081 GCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAAGCAGACAGAAAAATTC 1140
 QY
 1226 AGACGAAAAACAATAAAGAAAAAGAGAGGGACCAAAAGTTGGTTAAATTTATTCAGAG 1285
 DB
 1141 AGACGAAAAACAATAAAGAAAAAGAGAGGGACCAAAAGTTGGTTAAATTTATTCAGAG 1200

QY 1286 AATCAAGATTTGTAGATAAATTCATATGAAAGTACATCTTGTGTAACAAATAAATGC 1345
DB 1201 AATCAAGATTTGTAGATAAATTCATATGAAAGTACATCTTGTGTAACAAATAAATGC 1260
QY 1346 CACCCAGATCAAAACAAACACTTAGATTTCTCTGAAGGAAATTAATAATGTTGCTGTATTG 1405
DB 1261 CACCCAGATCAAAACAAACACTTAGATTTCTCTGAAGGAAATTAATAATGTTGCTGTATTG 1320
QY 1406 GAGTTTGATCTCTGAGTCTAAACATCAATGAGTGTCTAAAGCTTACAAAGAAAGCCGAGTA 1465
DB 1321 GAGTTTGATCTCTGAGTCTAAACATCAATGAGTGTCTAAAGCTTACAAAGAAAGCCGAGTA 1380
QY 1466 GCAAACTTCACTTTCCAAAGTGTATGATGAAACAGAAACCAACCAATGAGACGANT 1525
DB 1381 GCAAACTTCACTTTCCAAAGTGTATGATGAAACAGAAACCAACCAATGAGACGANT 1440
QY 1526 TCTACTCTAAATCTTTTACCACAA 1549
DB 1441 TCTACTCTAAATCTTTTACCACAA 1464

RESULT 13
ADB48211
ID ADB48211 standard; cDNA; 4771 BP.
XX AC ADB48211;
XX DT 04-DEC-2003 (first entry)
XX DE Novel human cDNA SEQ ID NO 121.
XX SS; cancer; neurodegenerative disease; human.
XX KW Homo sapiens.
XX OS US2003104529-A1.
XX PN 05-JUN-2003.
XX PD 04-JAN-2002; 2002US-00037270.
XX PF 21-JAN-2000; 2000US-00488725.
XX PR 25-APR-2000; 2000US-0052317.
XX PR 19-JUL-2000; 2000US-00620312.
XX XX (ZHOU/) ZHOU P.
XX PA (TANG/) TANG Y T.
XX PA (LIUC/) LIU C.
XX PA (ASUN/) ASUNDI V.
XX PA (DRNA/) DRMANAC R T.
XX XX Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX WPI; 2003-678194/64.
XX DR New polynucleotide, useful for treating diseases e.g., cancer or
XX PT neurodegenerative diseases.
XX PS Claim 1; SEQ ID NO 121; 99pp; English.
XX CC The invention relates to a polynucleotide comprising a sequence given in
XX CC the specification, or its mature protein-coding portion, or its
XX CC complement. The polynucleotide is useful for treating diseases e.g.,
XX CC cancer or neurodegenerative diseases and many others listed in the
XX CC specification. The present sequence represents a novel human cDNA. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format directly from USPTO
XX CC at seqdata.uspto.gov/sequence.html?DocID=20030104529.
XX SQ Sequence 4771 BP; 1740 A; 833 C; 909 G; 1289 T; 0 U; 0 Other;
Query Match 94.4%; Score 1462.4; DB 8; Length 4771;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 86 ATGGCAAGCAACTTAACTTCCAGAAATAACAGATGATTGCAAAAAGAGGATGTAAAT 145
DB 1 ATGGCAAGCAACTTAACTTCCAGAAATAACAGATGATTGCAAAAAGAGGATGTAAAT 60
QY 146 CAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTTTGACTCAACAAGAC 205
DB 61 CAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTTTGACTCAACAAGAC 120
QY 206 TGAATGAGAGCAGTCTTGAAGTGTGTTTAAAAAAGAACATCTTGTGTTGATATGGGCATCA 265
DB 121 GTGAATGAGAGCAGTCTTGAAGTGTGTTTAAAAAAGAACATCTTGTGTTGATATGGGCATCA 180
QY 266 CATGGACACAGCTATTCAAATAGAGAACTTATCAAAAGAAATTCGCGAAACACAGCCATTGAA 325
DB 181 CATGGACACAGCTATTCAAATAGAGAACTTATCAAAAGAAATTCGCGAAACACAGCCATTGAA 240
QY 326 GATTCGATTGAGACATCTAAGATGGGAAAGCCAGGTAAAAATGCTTCTTAAGACCAAACT 385
DB 241 GATTCGATTGAGACATCTAAGATGGGAAAGCCAGGTAAAAATGCTTCTTAAGACCAAACT 300
QY 386 GTGCTCTCAAAAGGAGGTAGAGAACTTCAAAAGCAAAACAAAGGTTAAAGAGACCCA 445
DB 301 GTGCTCTCAAAAGGAGGTAGAGAACTTCAAAAGCAAAACAAAGGTTAAAGAGACCCA 360
QY 446 GATATGGCTAAATCCGTCCTGCAATGAGTACAACTGCTTAAAGGTTCTAAGTCACTAAAGTT 505
DB 361 GATATGGCTAAATCCGTCCTGCAATGAGTACAACTGCTTAAAGGTTCTAAGTCACTAAAGTT 420
QY 506 GAGCTCATAGAGATTAATAATAGATTATACAAGGAAGGCAACCATCCATAGACCTGACA 565
DB 421 GAGCTCATAGAGATTAATAATAGATTATACAAGGAAGGCAACCATCCATAGACCTGACA 480
QY 566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTCAAGTTGGATTTTACT 625
DB 481 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTCAAGTTGGATTTTACT 540
QY 626 CTACAGCCTGAAACAGGACAGGCAATCTCATTTGATCCGATACATGAATTCAGGCTTC 585
DB 541 CTACAGCCTGAAACAGGACAGGCAATCTCATTTGATCCGATACATGAATTCAGGCTTC 600
QY 686 ACAATATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC 745
DB 601 ACAATATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTTC 660
QY 746 CGATTTGCTTCAGCTTGTATGAATTCACGTACAAATGGCACTATTTCATTTTGGAGTCAAA 805
DB 661 CGATTTGCTTCAGCTTGTATGAATTCACGTACCAATGGCACTATTTCATTTTGGAGTCAAA 720
QY 806 GACAAACCCCATGGGAAATTTGTTGGCATCAAGTCCCAATGATACCAAGGAGCCCTC 865
DB 721 GACAAACCCCATGGGAAATTTGTTGGCATCAAGTCCCAATGATACCAAGGAGCCCTC 780
QY 866 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCAATCAAGTCCCAAGCA 925
DB 781 ATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCAATCAAGTCCCAAGCA 840
QY 926 AAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAAAATAGTCTTATCT 985
DB 841 AAGAAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACTGCAAAATAGTCTTATCT 900
QY 986 GACAGATTTGTTTATGAAGTGGACATTTATCCACAGTCTCTCTGAATGCCAATATGATTAC 1045
DB 901 GACAGATTTGTTTATGAAGTGGACATTTATCCACAGTCTCTCTGAATGCCAATATGATTAT 960
QY 1046 TTCAGATTTAAATGCAAAATTTACAACACAAATATGGAAACAAAGTAAAAATTTCTCA 1105
DB 961 TTCAGATTTAAATGCAAAATTTACAACACAAATATGGAAACAAAGTAAAAATTTCTCA 1020
QY 1106 CTATTTGTGCGAGATGGGACCCAGCTCTAAGGACATTTACGAAAAATAAAGTTGATTTTCTCA 1165

Db	481	TGTTGATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGATTTAGT	540
Qy	626	CTACAGCCTGAAACAGGACCGAGCAATCTCATGATCCGATACATGAATTCAAAGCCTTC	685
Db	541	CTACAGCCTGAAACAGGACCGAGCAATCTCATGATCCGATACATGAATTCAAAGCCTTC	600
Qy	686	ACAAATACAGCAACAGCAGCAGAGAGGATGTCAGAGATGAATTTAGCAATGAGGTTTC	745
Db	601	ACAAATACAGCAACAGCAGCAGAGAGGATGTCAGAGATGAATTTAGCAATGAGGTTTC	660
Qy	746	CGATTTGCTTACGTTGTATGAATTCACGTAACAATGGCACTATTCTATTTGGAGTCAAA	805
Db	661	CGATTTGCTTACGTTGTATGAATTCACGTAACAATGGCACTATTCTATTTGGAGTCAAA	720
Qy	806	GACAAACCCCATGGGAAATTTGTGGCATCAAGTCAACCAATGATACCAAGGAGCCCTC	865
Db	721	GACAAACCCCATGGGAAATTTGTGGCATCAAGTCAACCAATGATACCAAGGAGCCCTC	780
Qy	866	ATTAACCATTTCAATCTGATGATAACAAGTATTTGAAGACCATCAAGTCCAAAGCA	925
Db	781	ATTAACCATTTCAATCTGATGATAACAAGTATTTGAAGACCATCAAGTCCAAAGCA	840
Qy	926	AGAAGTGATCTGAGAGCCAGATTTGTGGAGTTTACTGCCAAATAGTACTCTATCT	985
Db	841	AGAAGTGATCTGAGAGCCAGATTTGTGGAGTTTACTGCCAAATAGTACTCTATCT	900
Qy	986	GACAGATTTGTTATTTGAAGTGACATTAATTCACAGTCTCTGAAATGCAATATGATTAC	1045
Db	901	GACAGATTTGTTATTTGAAGTGACATTAATTCACAGTCTCTGAAATGCAATATGATTAC	960
Qy	1046	TTCCAGATTAATGCAATTAACACCAACAAATATGGCAAAAGTAAATAATCTCA	1105
Db	961	TTCCAGATTAATGCAATTAACACCAACAAATATGGCAAAAGTAAATAATCTCA	1020
Qy	1106	CTATTTGTGCGAGATGGGACCGCTCTAAGGACATTCAGAAAAATAAAGTTGATTCAGA	1165
Db	1021	CTATTTGTGCGAGATGGGACCGCTCTAAGGACATTCAGAAAAATAAAGTTGATTCAGA	1080
Qy	1166	GCATTTAAAGCAGATTTTAAACACCTGGCAGAGTCCAGAAAGCAGAGAAATTC	1225
Db	1081	GCATTTAAAGCAGATTTTAAACACCTGGCAGAGTCCAGAAAGCAGAGAAATTC	1140
Qy	1226	AGAGCAAAACAAATAAAGAGAGAGGAGGACCAAGTTGGTTAAATTTATTCACAGGA	1285
Db	1141	AGAGCAAAACAAATAAAGAGAGAGGAGGACCAAGTTGGTTAAATTTATTCACAGGA	1200
Qy	1286	AATCAAGATTTGTAGATTAATTCATATGATGATGATGATGATGATGATGATGATGATG	1345
Db	1201	AATCAAGATTTGTAGATTAATTCATATGATGATGATGATGATGATGATGATGATGATG	1260
Qy	1346	CACCCAGATCAAAACAAACACTTAGATTTCCCTGAAGGAAATTAATTTGCTGTATTG	1405
Db	1261	CACCCAGATCAAAACAAACACTTAGATTTCCCTGAAGGAAATTAATTTGCTGTATTG	1320
Qy	1406	GAGTTTGATCTGAGTCTAATCAATGAGTGGTCAAAAGCTTACAAAGAAAGCGAGTA	1465
Db	1321	GAGTTTGATCTGAGTCTAATCAATGAGTGGTCAAAAGCTTACAAAGAAAGCGAGTA	1380
Qy	1466	GCAAACTTCACCTTCCAGTGTATATGATGATGATGATGATGATGATGATGATGATGATG	1525
Db	1381	GCAAACTTCACCTTCCAGTGTATATGATGATGATGATGATGATGATGATGATGATGATG	1440
Qy	1526	TCTACTCTAAATCTTTACCATCAA	1549
Db	1441	TCTACTCTAAATCTTTACCATCAA	1464

RESULT 15

AAK77710

ID AAK77710 standard; DNA; 1045 BP.

XX

AC

AAK77710;

XX

DT	07-NOV-2001	(first entry)
XX	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32522.	
DE	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; cytostatic; gene therapy; vaccine; metastasis; ds.	
XX	Homo sapiens.	
OS	W0200157182-A2.	
PN	09-AUG-2001.	
PD	17-JAN-2001; 2001WO-US001354.	
PF	31-JAN-2000; 2000US-0179065P.	
XX	04-FEB-2000; 2000US-0180628P.	
PR	24-FEB-2000; 2000US-0184664P.	
PR	02-MAR-2000; 2000US-0186350P.	
PR	16-MAR-2000; 2000US-0189874P.	
PR	17-MAR-2000; 2000US-0190076P.	
PR	18-APR-2000; 2000US-0198123P.	
PR	19-MAY-2000; 2000US-0205515P.	
PR	07-JUN-2000; 2000US-0209467P.	
PR	28-JUN-2000; 2000US-0214866P.	
PR	30-JUN-2000; 2000US-0215135P.	
PR	07-JUL-2000; 2000US-0216647P.	
PR	07-JUL-2000; 2000US-0216880P.	
PR	11-JUL-2000; 2000US-0217487P.	
PR	11-JUL-2000; 2000US-0217496P.	
PR	14-JUL-2000; 2000US-0218290P.	
PR	26-JUL-2000; 2000US-0220963P.	
PR	26-JUL-2000; 2000US-0220964P.	
PR	14-AUG-2000; 2000US-0224518P.	
PR	14-AUG-2000; 2000US-0224519P.	
PR	14-AUG-2000; 2000US-0225213P.	
PR	14-AUG-2000; 2000US-0225214P.	
PR	14-AUG-2000; 2000US-0225266P.	
PR	14-AUG-2000; 2000US-0225267P.	
PR	14-AUG-2000; 2000US-0225268P.	
PR	14-AUG-2000; 2000US-0225270P.	
PR	14-AUG-2000; 2000US-0225447P.	
PR	14-AUG-2000; 2000US-0225757P.	
PR	14-AUG-2000; 2000US-0225758P.	
PR	14-AUG-2000; 2000US-0225759P.	
PR	18-AUG-2000; 2000US-0226279P.	
PR	22-AUG-2000; 2000US-0226681P.	
PR	22-AUG-2000; 2000US-0226688P.	
PR	22-AUG-2000; 2000US-0227182P.	
PR	23-AUG-2000; 2000US-0227009P.	
PR	30-AUG-2000; 2000US-0228924P.	
PR	01-SEP-2000; 2000US-0229287P.	
PR	01-SEP-2000; 2000US-0229343P.	
PR	01-SEP-2000; 2000US-0229344P.	
PR	01-SEP-2000; 2000US-0229345P.	
PR	05-SEP-2000; 2000US-0229509P.	
PR	05-SEP-2000; 2000US-0229513P.	
PR	06-SEP-2000; 2000US-0230437P.	
PR	06-SEP-2000; 2000US-0230438P.	
PR	08-SEP-2000; 2000US-0231242P.	
PR	08-SEP-2000; 2000US-0231243P.	
PR	08-SEP-2000; 2000US-0231244P.	
PR	08-SEP-2000; 2000US-0231413P.	
PR	08-SEP-2000; 2000US-0231414P.	
PR	08-SEP-2000; 2000US-0232080P.	
PR	12-SEP-2000; 2000US-0232081P.	
PR	12-SEP-2000; 2000US-0231968P.	
PR	14-SEP-2000; 2000US-0232397P.	
PR	14-SEP-2000; 2000US-0232398P.	
PR	14-SEP-2000; 2000US-0232399P.	
PR	14-SEP-2000; 2000US-0232400P.	
PR	14-SEP-2000; 2000US-0232401P.	
PR	14-SEP-2000; 2000US-0233063P.	

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-02411221P.
PR 20-OCT-2000; 2000US-02411785P.
PR 20-OCT-2000; 2000US-02411786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251900P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 32522; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK64702 to AAK64702. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK67694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK62169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 1045 BP; 410 A; 187 C; 196 G; 252 T; 0 U; 0 Other;
Query Match 67.4%; Score 1043.4; DB 4; Length 1045;
Best Local Similarity 99.9%; Pred. No. 2.6e-223;
Matches 1044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 206 GTGAATGGAGCAGCTTGAAGTGGTTAAAGAAAGCAATCTTGTGATATGGGCATCACA 265
Db 1 GTGAATGGAGCAGCTTGAAGTGGTTAAAGAAAGCAATCTTGTGATATGGGCATCACA 60
QY 266 CATGGACCAGCTATTCAAAATAGAGAACTATTCAAGAAATTCGCGAAACAGCCATTGAA 325
Db 61 CATGGACCAGCTATTCAAAATAGAGAACTATTCAAGAAATTCGCGAAACAGCCATTGAA 120
QY 326 GATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAGACCAACT 385
Db 121 GATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAGACCAACT 180
QY 386 GTGTCTCAAAAGGACGTAGAGAACTTCAAGCAAAACAAAGGGTAAAGAGAACCCA 445
Db 181 GTGTCTCAAAAGGACGTAGAGAACTTCAAGCAAAACAAAGGGTAAAGAGAACCCA 240
QY 446 GATATGGCTAATCCGCTGCAATGAGTACAACTGCTAAAGTTTCTAAGTCTCTAAAGTT 505
Db 241 GATATGGCTAATCCGCTGCAATGAGTACAACTGCTAAAGTTTCTAAGTCTCTAAAGTT 300
QY 506 GAGCTCATAGAGATATAATAGATTATCAAGAAAGGCAACCATCCATAGACCTGACA 565
Db 301 GAGCTCATAGAGATATAATAGATTATCAAGAAAGGCAACCATCCATAGACCTGACA 360
QY 566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTCAAGTTGGATTTAGT 625
Db 361 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTTCAAGTTGGATTTAGT 420
QY 626 CTACAGCCTGAAACAGGACCGCAATCTCATTCGATACATGATTAATCCAAAGCCTTC 685
Db 421 CTACAGCCTGAAACAGGACCGCAATCTCATTCGATACATGATTAATCCAAAGCCTTC 480

QY	586	ACAAATACAGCACAGCCACAGAGAGGATGTCAAGATGAATTTAGCAATGAGGTTTTC	745
Db	481	ACAAATACAGCACAGCCACAGAGAGGATGTCAAGATGAATTTAGCAATGAGGTTTTC	540
QY	746	CGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTGGAGTCAAA	805
Db	541	CGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTGGAGTCAAA	600
QY	806	GACAAACCCCATGGGAAATTTGTCGATCAAGTCCCAATGATATACCAAGGAGCCCTC	865
Db	601	GACAAACCCCATGGGAAATTTGTCGATCAAGTCCCAATGATATACCAAGGAGCCCTC	660
QY	866	ATTAAACCATTTCAATCTGATGATATAACAAGTATTTTGAAGACCATCAAGTCCCAACAGCA	925
Db	661	ATTAAACCATTTCAATCTGATGATATAACAAGTATTTTGAAGACCATCAAGTCCCAACAGCA	720
QY	926	AAGAAGTGCATTCGAGGCCAAGATTTGTGGAAGTTTACTGCCAATAGTACTCTATCT	985
Db	721	AAGAAGTGCATTCGAGGCCAAGATTTGTGGAAGTTTACTGCCAATAGTACTCTATCT	780
QY	986	GACAGATTTGTTATTTGAAGTGGACATTTATCCACAGTTCTCTGAATGCCAATATGATTAC	1045
Db	781	GACAGATTTGTTATTTGAAGTGGACATTTATCCACAGTTCTCTGAATGCCAATATGATTAT	840
QY	1046	TTCCAGATTAATAATGCAAAATTAACAACAATATGGGACAAAGTAAATAATTTCTCA	1105
Db	841	TTCCAGATTAATAATGCAAAATTAACAACAATATGGGACAAAGTAAATAATTTCTCA	900
QY	1106	CTATTTGTGCGAGATGGACAGCTCTAAGGACATTAAGGACATTAAGGTTGATTTCAGA	1165
Db	901	CTATTTGTGCGAGATGGACAGCTCTAAGGACATTAAGGACATTAAGGTTGATTTCAGA	960
QY	1166	GCATTTAAAGCAGATTTTAAACACTGCGCAGATCCAGAAAGCAGCAGAGAAATTC	1225
Db	961	GCATTTAAAGCAGATTTTAAACACTGCGCAGATCCAGAAAGCAGCAGAGAAATTC	1020
QY	1226	AGAGCAAAACAAATATAAAAGAAA	1250
Db	1021	AGAGCAAAACAAATATAAAAGAAA	1045

Search completed: April 25, 2004, 08:46:12
Job time : 417.576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 4078.84 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-17
Perfect score: 1549
Sequence: 1 aacacatcagatatttcag.....ctctaaatcttaccatcaa 1549

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :					GenEmbl :				
1:	gb	ba	*		1:	gb	ba	*	
2:	gb	hg	*		2:	gb	hg	*	
3:	gb	in	*		3:	gb	in	*	
4:	gb	om	*		4:	gb	om	*	
5:	gb	ov	*		5:	gb	ov	*	
6:	gb	pat	*		6:	gb	pat	*	
7:	gb	ph	*		7:	gb	ph	*	
8:	gb	pl	*		8:	gb	pl	*	
9:	gb	pr	*		9:	gb	pr	*	
10:	gb	ro	*		10:	gb	ro	*	
11:	gb	stb	*		11:	gb	stb	*	
12:	gb	sy	*		12:	gb	sy	*	
13:	gb	un	*		13:	gb	un	*	
14:	gb	vi	*		14:	gb	vi	*	
15:	em	ba	*		15:	em	ba	*	
16:	em	fun	*		16:	em	fun	*	
17:	em	in	*		17:	em	in	*	
18:	em	in	*		18:	em	in	*	
19:	em	mu	*		19:	em	mu	*	
20:	em	om	*		20:	em	om	*	
21:	em	or	*		21:	em	or	*	
22:	em	ov	*		22:	em	ov	*	
23:	em	pat	*		23:	em	pat	*	
24:	em	ph	*		24:	em	ph	*	
25:	em	pl	*		25:	em	pl	*	
26:	em	ro	*		26:	em	ro	*	
27:	em	stb	*		27:	em	stb	*	
28:	em	un	*		28:	em	un	*	
29:	em	vi	*		29:	em	vi	*	
30:	em	hg	hum	*	30:	em	hg	hum	*
31:	em	hg	inv	*	31:	em	hg	inv	*
32:	em	hg	other	*	32:	em	hg	other	*
33:	em	hg	mus	*	33:	em	hg	mus	*
34:	em	hg	pln	*	34:	em	hg	pln	*
35:	em	hg	rod	*	35:	em	hg	rod	*
36:	em	hg	nam	*	36:	em	hg	nam	*
37:	em	hg	vrt	*	37:	em	hg	vrt	*
38:	em	sy	*		38:	em	sy	*	
39:	em	hgco	hum	*	39:	em	hgco	hum	*
40:	em	hgco	mus	*	40:	em	hgco	mus	*
41:	em	hgco	other	*	41:	em	hgco	other	*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1533.8	99.0	3956	9	AK122951	AK122951 Homo sapi
2	1472.4	95.1	6826	9	HSM808722	BX647072 Homo sapi
3	1472.4	95.1	6827	9	AB095925	AB095925 Homo sapi
C 4	1472	95.0	191191	9	AC000119	AC000119 Homo sapi
C 5	1464	94.5	172371	2	AC092871	AC092871 Pan trogl
C 6	1462.4	94.4	4771	6	AR338630	AR338630 Sequence
7	1462.4	94.4	6233	6	AC098228	AC098228 Sequence
C 8	1396.8	90.2	176350	2	AC092513	AC092513 Papio anu
C 9	1048.6	67.7	175424	2	AC105435	AC105435 Bos scrof
C 10	1039.8	67.1	108538	2	AC108893	AC108893 Bos tauru
C 11	921	59.5	148278	2	AC079885	AC079885 Rattus no
C 12	921	59.5	180665	2	AC084240	AC084240 Rattus no
C 13	921	59.5	221557	2	AC134055	AC134055 Rattus no
C 14	788.4	50.9	119266	2	AC108900	AC108900 Fells cat
C 15	759	49.0	225885	2	AC113236	AC113236 Canis fam
C 16	672	43.4	705	6	AB095926	AB095926 Homo sapi
17	636.8	41.1	7122	9	AB095926	AB095926 Sequence
18	628.4	40.6	5870	6	AX098226	AX098226 Canis fam
C 19	599.6	38.7	137838	2	AC110666	AC110666 Mus muscu
C 20	499	32.2	188579	2	AC107316	AC107316 Mus muscu
C 21	498	32.1	195282	2	AC022453	AC022453 Rattus no
C 22	473.8	30.6	125962	2	AC084271	AC084271 Rattus no
C 23	473.8	30.6	191284	2	AC129122	AC129122 Rattus no
24	469	30.3	262697	2	AC108642	AC108642 Rattus no
25	314.6	20.3	52652	2	AC104854	AC104854 Homo sapi
26	169.2	10.9	188729	2	AX005331	AX005331 Danio rer
27	138	8.9	356	6	AX245250	AX245250 Sequence
C 28	97.8	6.3	262897	2	AC108642	AC108642 Rattus no
C 29	95.2	6.1	147824	2	EX548026	EX548026 Danio rer
C 30	95.2	6.1	157603	5	AL953865	AL953865 Zebrafish
31	73	4.7	251	6	AX246569	AX246569 Sequence
C 32	72.2	4.7	1453	8	AJ591978	AJ591978 Arabidops
C 33	69.6	4.5	1434	8	AJ592058	AJ592058 Arabidops
C 34	69.4	4.5	139827	5	AL929509	AL929509 Zebrafish
C 35	68.6	4.4	7218	6	I66494	I66494 Sequence 14
C 36	68.2	4.4	93791	2	AC138073	AC138073 Homo sapi
37	67.8	4.4	188355	2	BX537350	BX537350 Danio rer
38	66.4	4.3	347050	3	PFA929351	AL929351 Plasmodiu
C 39	66.2	4.3	60604	2	AC023466	AC023466 Homo sapi
C 40	66	4.3	1407	8	AJ592026	AJ592026 Arabidops
C 41	63	4.1	188477	2	BX470139	BX470139 Danio rer
C 42	61.6	4.0	349980	6	AX344567	AX344567 Sequence
C 43	61	3.9	14006	6	AX346860	AX346860 Sequence
C 44	60.8	3.9	128683	2	AC135793	AC135793 Oryza sat
C 45	60.8	3.9	143585	2	AC013349	AC013349 Homo sapi

ALIGNMENTS

RESULT 1
AK122951
LOCUS Homo sapiens cDNA FLJ16668 fis, clone THYMU2039788. 3956 bp mRNA linear PRI 09-SEP-2003
DEFINITION
ACCESSION AK122951
VERSION AK122951.1 GI:34528396
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Inoue, N., Takeuchi, K.,
Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J.,

Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahari, K., Masuno, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3956)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUN-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 252-0818, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..3956
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYM2039788"
/tissue_type="thymus"
/clone_lib="THYM2"
/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 99.0%; Score 1533.8; DB 9; Length 3956;
Best Local Similarity 99.8%; Pred. No. 1.2e-279;
Matches 1546; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
1 AACACATCAGATATTTTCAGCACTAAAGAGATGGTTTCCCAATATATATAAGAA 60
24 AACACATCAGATATTTTCAGCACTAAAGAGATGGTTTCCCAATATATATAAGAA 83
61 ATTTCAGCACTACTGGGTATCAGATGCGAAGCACTTAACTTCCAGAAATACAGA 120
84 ATTTCAGCACTACTGGGTATCAGATGCGAAGCACTTAACTTCCAGAAATACAGA 143
121 TGATTGGCAAAAAGAGATGTAATCAGTGGTTAGAAAGTCTAAGATGACCAAAAACA 180
144 TGATTGGCAAAAAGAGATGTAATCAGTGGTTAGAAAGTCTAAGATGACCAAAAACA 203
181 CAGGGAATTTTCACTGAACAGACGCTGAATGCGAGCAGCTTGAAGTGGTTAAAAAAGA 240
204 CAGGGAATTTTGACTGAACAGACGCTGAATGCGAGCAGCTTGAAGTGGTTAAAAAAGA 263
241 ACATCTTGTGATATGGCATCAACATGACCGACGCTATTCAAAATAGAAAGAACTATTCAA 300
264 ACATCTTGTGATATGGCATCAACATGACCGACGCTATTCAAAATAGAAAGAACTATTCAA 323
301 AGAATTCGGAAAAACGCCATTGAAGATTCGATTCAGACATCAAGATGGGAAGCCGAG 360
324 AGAATTCGGAAAAACGCCATTGAAGATTCGATTCAGACATCAAGATGGGAAGCCGAG 383
361 TAAAAATGCTCCCTAAAGACCAAACTGTGCTCAAAAAGGAACGTAGAGAACTTCAAAGCA 420
384 TAAAAATGCTCCCTAAAGACCAAACTGTGCTCAAAAAGGAACGTAGAGAACTTCAAAGCA 443
421 AAAACAAAGGGTAAAGAGAACCCAGATATGGCTAATTCGCTGCAATGAGTACAACTGC 480
444 AAAACAAAGGGTAAAGAGAACCCAGATATGGCTAATTCGCTGCAATGAGTACAACTGC 503
481 TAAAGGTTCTAAGTCACTAAAAAGTTGAGCTCATAGAGATAAAATAGATTATACAAAGGA 540
504 TAAAGGTTCTAAGTCACTAAAAAGTTGAGCTCATAGAGATAAAATAGATTATACAAAGGA 563
541 AAGGCAACCATCCATAGACCTGATGTGATCATATATCCATTTGATGAAATCAGTAATCC 600

564 AAGGCAACCATCCATAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCC 623
601 ATATCGTTACAGTTTGGATTTAGTCTACAGCTGGAACAGGACCGAGGAATCTCATTTGA 660
624 ATATCGTTACAGTTTGGATTTAGTCTACAGCTGGAACAGGACCGAGGAATCTCATTTGA 683
661 TCCGATACATGAATTCAAAGCCCTTCAAAATACAGCAACAGCCACAGAGAGATGTCAA 720
684 TCGATACATGAATTCAAAGCCCTTCAAAATACAGCAACAGCCACAGAGAGATGTCAA 742
721 GATGAATTTAGCAATGAGGTTTCCGATTTGCTGAGCTTGTATGAATTCAGTACCAA 780
743 GATGAATTTAGCAATGAGGTTTCCGATTTGCTGAGCTTGTATGAATTCAGTACCAA 802
781 TGGCACTATTTCATTTGGAGTCAAGCAACACCCCATGGGAAAATTTGGTCATCAAGT 840
803 TGGCACTATTTCATTTGGAGTCAAGCAACACCCCATGGGAAAATTTGGTCATCAAGT 862
841 CACCAATGATACCAAGGAGCCCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTT 900
863 CACCAATGATACCAAGGAGCCCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTT 922
901 TGAAGACCATCAAGTCCAAACAGCAAGAGAGTGCATTCGAGAGCCCAAGATTTTGTGAAGT 960
923 TGAAGACCATCAAGTCCAAACAGCAAGAGAGTGCATTCGAGAGCCCAAGATTTTGTGAAGT 982
961 TTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTTATGAAGTGGACATTTATTCACA 1020
983 TTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTTATGAAGTGGACATTTATTCACA 1042
1021 GTTCTCTGAATGCCAATATGATTACTTCCAGATTTAAATGCAAAAATTTACAACAACAAAT 1080
1043 GTTCTCTGAATGCCAATATGATTATTTCCAGATTTAAATGCAAAAATTTACAACAACAAAT 1102
1081 ATGGGAACAAGTAAAAAATTTCTCACTATTTTGCAGATGGGACCGAGCTCTAAGGACAT 1140
1103 ATGGGAACAAGTAAAAAATTTCTCACTATTTTGCAGATGGGACCGAGCTCTAAGGACAT 1162
1141 TAGCAAAAATAAAGTTGATTTTCAGAGCATTTAAAGCAGATTTTAAACAACCTGCGAGATC 1200
1163 TAGCAAAAATAAAGTTGATTTTCAGAGCATTTAAAGCAGATTTTAAACAACCTGCGAGATC 1222
1201 CAGAAAAGCAGAGAGAAAATTTTCAGCAAAAACAAATAAATAAAGAGAGAGAGGACC 1260
1223 CAGAAAAGCAGAGAGAAAATTTTCAGCAAAAACAAATAAATAAAGAGAGAGAGGACC 1282
1261 AAGCTGGTTAAATTTATGACAGCAATCAAGATTTGTTAGATAATTCATCATATGAACA 1320
1283 AAGCTGGTTAAATTTATGACAGCAATCAAGATTTGTTAGATAATTCATCATATGAACA 1342
1321 GTACATTTCTTTGTAACAAATAAATGCCACCCAGATCAAAACAAAACACTAGATTTCTGAA 1380
1343 GTACATTTCTTTGTAACAAATAAATGCCACCCAGATCAAAACAAAACACTAGATTTCTGAA 1402
1381 GGAATTTAAATGGTTCTGCTGTTTGGAGTTTGAATCTGAGTCTAAACATCAATGGAGTGT 1440
1403 GGAATTTAAATGGTTCTGCTGTTTGGAGTTTGAATCTGAGTCTAAACATCAATGGAGTGT 1462
1441 CAAAGCTTACAAAGAAAGCCGAGTAGCAAACTTCATCTTTCCAAAGTGTATATGTAGAAC 1500
1463 CAAAGCTTACAAAGAAAGCCGAGTAGCAAACTTCATCTTTCCAAAGTGTATATGTAGAAC 1522
1501 GAAACACACACCAATGAGAGGATTTCTACTCTAAATCTTTTACCATCAA 1549
1523 GAAACACACACCAATGAGAGGATTTCTACTCTAAATCTTTTACCATCAA 1571

RESULT 2

HSN808722

LOCUS HSN808722

DEFINITION Homo sapiens mRNA; cDNA DKFZp686B0866 (from clone DKFZp686B0866).

ACCESSION BX647072

6826 bp mRNA linear

PRI 30-AUG-2003

970 AAATAGTACTCTATCTGACAGATTGTTTATTGAAGTGACATTATTCACAGTTCTCTGA 1029
Db 93620 AAATAGTACTCTATCTGACAGATTGTTTATTGAAGTGACATTATTCACAGTTCTCTGA 93561
Qy 1030 ATGCCAATATGATTACTTCCAGATTAAATCAAAATACAAACAATAATATGGGAACA 1089
Db 93560 ATGCCAATATGATTACTTCCAGATTAAATCAAAATACAAACAATAATATGGGAACA 93501
Qy 1090 AAGTAAAAAATCTCATTATTTGTCGAGATGGGACAGCTCTTAAGGACATATTACGAAAAA 1149
Db 93500 AAGTAAAAAATCTCATTATTTGTCGAGATGGGACAGCTCTTAAGGACATATTACGAAAAA 93441
Qy 1150 TAAAGTTGATTTCAGAGCAATTTAAGCAGATTTTAAACACCTGCAGAGTCCAGAAAGC 1209
Db 93440 TAAAGTTGATTTCAGAGCAATTTAAGCAGATTTTAAACACCTGCAGAGTCCAGAAAGC 93381
Qy 1210 AGCAGAAGAAAAATTCAGAGCAAAAAAACAATAAAAAAGAGAGAGGACCAAAATGTTGGT 1269
Db 93380 AGCAGAAGAAAAATTCAGAGCAAAAAAACAATAAAAAAGAGAGAGGACCAAAATGTTGGT 93321
Qy 1270 TAAATATTGACAGAAATCAAGATTTGTTAGATTAATTCATATGACAGTACATCTCT 1329
Db 93320 TAAATATTGACAGAAATCAAGATTTGTTAGATTAATTCATATGACAGTACATCTCT 93261
Qy 1330 TGTAACAATAATGCCACCCAGATCAACAAACACCTTAGATTTCTGAAGGAAATTA 1389
Db 93260 TGTAACAATAATGCCACCCAGATCAACAAACACCTTAGATTTCTGAAGGAAATTA 93201
Qy 1390 ATGGTTTGCTGATTTGGAGTTTGATCTCGAGTCTTAACATCAATGAGTGTTCAAGCTTA 1449
Db 93200 ATGGTTTGCTGATTTGGAGTTTGATCTCGAGTCTTAACATCAATGAGTGTTCAAGCTTA 93141
Qy 1450 CAAAGAAAGCCAGTAGCAAACTTCACTTCCAGTGTATATGTAAGCAAGAAACCAAC 1509
Db 93140 CAAAGAAAGCCAGTAGCAAACTTCACTTCCAGTGTATATGTAAGCAAGAAACCAAC 93081
Qy 1510 ACCAATGAGAGATTCTTACTCTAAATCTTTTACCATCAA 1549
Db 93080 ACCAATGAGAGATTCTTACTCTAAATCTTTTACCATCAA 93041

RESULT 5
AC092871/c
LOCUS
DEFINITION Pan troglodytes clone RP43-120E8, WORKING DRAFT SEQUENCE, 3 ordered pieces.
AC092871 172371 bp DNA linear HTG 12-JUN-2002
AC092871.2 GI:21392483
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE Pan troglodytes
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1 (bases 1 to 172371)
AUTHORS Akheri, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blackley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., McGuire, C., Pearson, R., Portnoy, M.E., Prasad, A., Schueler, M.G., Stantropop, S., Thomas, J.W., Wetherby, K.D., Wiggins, L., Young, A., Vogt, J.L., Walker, M.A., Thomas, P.J., Touchman, J.W., Tsurgan, C., Vogt, J.L., Walker, M.A., and Green, E.D.

TITLE NISC Comparative Sequencing Initiative
REFERENCE 2 (bases 1 to 172371)
JOURNAL Unpublished
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717
REFERENCE 3 (bases 1 to 172371)
JOURNAL Grovemont Circle, Gaithersburg, MD 20877, USA

AUTHORS
TITLE
JOURNAL
COMMENT

Green, E.D.
Direct Submission
Submitted (12-JUN-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Jun 12, 2002 this sequence version replaced gi:15055304.
----- Genom Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center Project name: ank
Center Clone name: 120E08

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171808 bases at least Q40
Consensus quality: 172035 bases at least Q30
Consensus quality: 172123 bases at least Q20
Insert size: 129000; agarose-fp
Insert size: 137000; pulse-field-gel
Quality coverage: 15.17% in Q20 bases; agarose-fp
Quality coverage: 14.25% in Q20 bases; pulse-field-gel
Quality coverage: 11.37% in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 77640: contig of 77640 bp in length
* 77641 77740: gap of unknown length
* 77741 102026: contig of 24286 bp in length
* 102027 102126: gap of unknown length
* 102127 172371: contig of 70245 bp in length.

FEATURES

source
1. 172371
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-120E8"
/clone_lib="RP43"
1. 77640
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature
1. 33202
/notes="clone overlaps with GenBank Accession Number AC098646 clone RP43-44G6 (center project name anj)"
77741..102026
/notes="assembly_fragment"
102127..172371
/notes="assembly_fragment"
clone_end:T7

```
ORIGIN
Query Match 94.5%; Score 1464; DB 2; Length 172371;
Best Local Similarity 99.3%; Pred. No. 8.4e-267;
Matches 1470; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 70 ACTACTGGGTATCAGATGCGCAAGCAACTTAACCTTCAGAAATACAGATGATGGAC 129
DB 117802 ACATATAGGTATCAGATGCGCAAGCAACTTAACCTTCAGAAATACAGATGATGGAC 117743
QY 130 AAAAGAGGATGTAATACAGTGGTTAGAAAGTCAATAAGATTGACCAAAAAACACAGGGAAT 189
DB 117742 AAAAGAGGATGTAATACAGTGGTTAGAAAGTCAATAAGATTGACCAAAAAACACAGGGAAT 117683
QY 190 TTTGACTGAACAAGAGCGTAAGTGGAGCGTCTTGAAGTGGTTAAAGAAAGACATCTTGT 249
DB 117682 TTTGACTGAACAAGAGCGTAAGTGGAGCGTCTTGAAGTGGTTAAAGAAAGACATCTTGT 117623
QY 250 TGATATGGGCATCACATCGACCCAGCTATTCAATAGAGAACTATTCAAGAGAAATGGG 309
DB 117622 TGATATGGGCATCACATCGACCCAGCTATTCAATAGAGAACTATTCAAGAGAAATGGG 117563
QY 310 GAAAAAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAGATGC 369
DB 117562 GAAAAAGCCATTGAAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAAGATGC 117503
QY 370 TCCTAAAGACCAACTGTCTCAAGAGGACGTAGAGAACTTCAAGCAAAAAACAAA 429
DB 117502 TCCTAAAGACCAACTGTCTCAAGAGGACGTAGAGAACTTCAAGCAAAAAACAAA 117443
QY 430 GGGTAAAGAGAACCCAGATATGCTTAATCCGTCTGCAATGAGTACAACTGCTAAAGTTTC 489
DB 117442 GGGTAAAGAGAACCCAGATATGCTTAATCCGTCTGCAATGAGTACAACTGCTAAAGTTTC 117383
QY 490 TAACTCACTAAAGTTCAGCTCATAGAGATAAATAGATTATACAAAGGAAGGCAACC 549
DB 117382 TAACTCACTAAAGTTCAGCTCATAGAGATAAATAGATTATACAAAGGAAGGCAACT 117323
QY 550 ATCCATAGACCTGACATGTGTATCATATCCATTGTATGATTAATCAGTAATCCATATCGTTA 609
DB 117322 ATCCATAGACCTGACATGTGTATCATATCCATTGTATGATTAATCAGTAATCCATATCGTTA 117263
QY 610 CAAGTTGGATTTAGTCTACAGCTGAAACAGGACCGCAACTCTATTGATCCGATACA 669
DB 117262 CAAGTTGGATTTAGTCTACAGCTGAAACAGGACCGCAACTCTATTGATCCGATACA 117203
QY 670 TGAATTCAGAGCCCTTCACAAATACAGCAACAGCCACAGCAAGAGAGGTGTCAAGATGAAAT 729
DB 117202 TGAATTCAGAGCCCTTCACAAATACAGCAACAGCCACAGCAAGAGAGGTGTCAAGATGAAAT 117143
QY 730 TAGCAATGAGTTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTAT 789
DB 117142 TAGCAATGAGTTTTCGATTTGCTTCAGCTTGTATGAATTCAGTACCAATGGCACTAT 117083
QY 790 TCATTTTGGAGTCAAGACCAAAACCCATGGGAAAATTTGTTGGCATCAAAAGTCAACCAATGA 849
DB 117082 TCATTTTGGAGTCAAGACCAAAACCCATGGGAAAATTTGTTGGCATCAAAAGTCAACCAATGA 117023
QY 850 TACCAAGAGCCCTTCATTAACCAATTCATCTGATGATTAACAGATATTTTGAAGACCA 909
DB 117022 TACCAAGAGCCCTTCATTAACCAATTCATCTGATGATTAACAGATATTTTGAAGACCA 116963
QY 910 TCAAGTCCAAACAGCAAGAGTGCATTCGAGAGCCCAAGATTTGTGGAAGTTTACTTGC 969
DB 116962 TCAAGTCCAAACAGCAAGAGTGCATTCGAGAGCCCAAGATTTGTGGAAGTTTACTTGC 116903
QY 970 AAATAGTACTCTATCTGACAGATTTGTTATGAAAGTGGACATTAATTCACAGTTCTCTGA 1029
DB 116902 AAATAGTACTCTATCTGACAGATTTGTTATGAAAGTGGACATTAATTCACAGTTCTCTGA 116843
QY 1030 ATGCCAATATGATTTCCAGATTAATAATGCAAAATTAACAACAAAAATATGGGAACA 1089
DB 116842 ATGCCAATATGATTTCCAGATTAATAATGCAAAATTAACAACAAAAATATGGGAACA 116783
QY 1090 AAGTAAAAAATCTCACTATTTCGAGATGGGACCCAGCTCTAAGACANTTACGAAAAA 1149
DB 116782 AAGTAAAAAATCTCACTATTTCGAGATGGGACCCAGCTCTAAGGACANTTACGAAAAA 116723
QY 1150 TAAAGTTTGAATTCAGAGCATTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAAGC 1209
DB 116722 TAAAGTTTGAATTCAGAGCATTTAAAGCAGATTTTAAAAACACTGGCAGAGTCCAGAAAAAGC 116663
QY 1210 AGCAGAGAAAAAATTCAGAGCAAAAAACAATAAAAAAGAAAGAGAGGACCAAGTTGGT 1269
DB 116662 AGCAGAGAAAAAATTCAGAGCAAAAAACAATAAAAAAGAAAGAGAGGACCAAGTTGGT 116603
QY 1270 TAAATTTATTCAGAGCAAAATCAAGATTTGTTAGATAAATTCATATCACTATCAAGTACATTCT 1329
DB 116602 TAAATTTATTCAGAGCAAAATCAAGATTTGTTAGATAAATTCATATCACTATCAAGTACATTCT 116543
QY 1330 TGTAAACAAATAAATGCCACCCAGATCAAAACAAACACTTTAGATTTCCTGAAGAAATTA 1389
DB 116542 TGTAAACAAATAAATGCCACCCAGATCAAAACAAACACTTTAGATTTCCTGAAGAAATTA 116483
QY 1390 ATGGTTTCTGTTATTCGAGTTTCTGATCCTGAGTCTTAACATCAATGGAGTGTCAAAGCTTA 1449
DB 116482 ATGGTTTCTGTTATTCGAGTTTCTGATCCTGAGTCTTAACATCAATGGAGTGTCAAAGCTTA 116423
QY 1450 CAAAGAAAGCCGAGTACGCAACCTTCACTTCCAAAGTGTATATGTAGAACAGAAAAACAC 1509
DB 116422 CAAAGAAAGCCGAGTACGCAACCTTCACTTCCAAAGTGTATATGTAGAACAGAAAAACAC 116363
QY 1510 ACCAATGAGAGCGATTTCTACTCTAATCTTTTACCATCAA 1549
DB 116362 ACCAATGAGAGCGATTTCTACTCTAATCTTTTACCATCAA 116323
```

RESULT 6

AR338630

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Query Match

Best Local Similarity

Matches 1463; Conservative

QY 86

DB 1

QY 146

DB 61

QY 206

DB 121

QY 266

DB 325

AR338630

Sequence 121 from patent US 6569662.

AR338630.1 GI:33725487

Unknown.

Unclassified.

1 (bases 1 to 4771)

Tang, F.T., Zhou, P. and Drmanac, R.T.

Nucleic acids and polypeptides

Patent: US 6569662-A 121 27-MAY-2003;

Location/Qualifiers

1. 4771

/organism="unknown"

/mol_type="genomic DNA"

Query Match

Best Local Similarity

Matches 1463; Conservative

QY 86

DB 1

QY 146

DB 61

QY 206

DB 121

QY 266

DB 325

QY 146

DB 61

QY 206

DB 121

QY 266

DB 325

QY 146

DB 61

QY 206

DB 121

QY 266

DB 325

QY 146

DB 61

QY 206

DB 121

QY 266

DB 325

181 CATGGACCACTATTCAAATAGAGAACTATTCAAGAAATTTGGGAAAACAGCCATTGAA 240
326 GATTTCGATTCAGACATCTAAGATGGGAAAGCCAGTAATAAATCTCTCTTAAGACCAAACT 385
241 GATTTCGATTCAGACATCTAAGATGGGAAAGCCAGTAATAAATCTCTCTTAAGACCAAACT 300
386 GTGCTCTCAAAAGGAAGCTAGAGAACTTTCAAGCAAAACAAAGGGTAAAGAGAACCCCA 445
301 GTGCTCTCAAAAGGAAGCTAGAGAACTTTCAAGCAAAACAAAGGGTAAAGAGAACCCCA 360
446 GATATGCTAATCCGTCGCAATGAGTACAACTGCTTAAAGGTTCTAAGTCACCTAAAGTT 505
361 GATATGCTAATCCGTCGCAATGAGTACAACTGCTTAAAGGTTCTAAGTCACCTAAAGTT 420
506 GAGCTCATAGAGATAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTTGACA 565
421 GAGCTCATAGAGATAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTTGACA 480
566 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGGATTTAGT 625
481 TGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAAAGTTGGATTTAGT 540
626 CTACAGCTGAAACAGGACGAGCAATCTCAATGATCGGATACATGATGAATTCAAAGCTTC 685
541 CTACAGCTGAAACAGGACGAGCAATCTCAATGATCGGATACATGATGAATTCAAAGCTTC 600
686 ACAATATCAGCAACAGCCACACAGAGAGAGATGTCAGATGAATTTAGCAATGAGTTTC 745
601 ACAATATCAGCAACAGCCACACAGAGAGAGATGTCAGATGAATTTAGCAATGAGTTTC 660
746 CGATTTGCTCAGCTTGATGAATTCAGTACCAATGCGACTATCTATCTTTGGATGCAAA 805
661 CGATTTGCTCAGCTTGATGAATTCAGTACCAATGCGACTATCTATCTTTGGATGCAAA 720
806 GACAAACCCCATGGGAAATTTGTTGGCATCAAGTCCCAATGATGATCAAGGAGCCCTC 865
721 GACAAACCCCATGGGAAATTTGTTGGCATCAAGTCCCAATGATGATCAAGGAGCCCTC 780
866 ATTAACCAATTCATCTGATGATGAATTCAGTACCAATGCGACTATCTATCTTTGGATGCAAA 925
781 ATTAACCAATTCATCTGATGATGAATTCAGTACCAATGCGACTATCTATCTTTGGATGCAAA 840
926 AAGAGTGCATTCGAGACCCAGATTTGTTGGAGTTTACTGCGCAATAGTACTCTATCT 985
841 AAGAGTGCATTCGAGACCCAGATTTGTTGGAGTTTACTGCGCAATAGTACTCTATCT 900
986 GACAGATTTGTTATGAAATGAGATTAATTCAGATTTCTGATGATGCAATGATGATAC 1045
901 GACAGATTTGTTATGAAATGAGATTAATTCAGATTTCTGATGATGCAATGATGATAC 960
1046 TTCCAGATTAATGCAAAATTTACAAACAAATATCGGAAACAAAGTAAATAATTTCTCA 1105
961 TTCCAGATTAATGCAAAATTTACAAACAAATATCGGAAACAAAGTAAATAATTTCTCA 1020
1106 CTATTTGTCGAGATGGGACCACTGCTTAAGGACATTAAGGAAATTAAGTTGATTCAGA 1165
1021 CTATTTGTCGAGATGGGACCACTGCTTAAGGACATTAAGGAAATTAAGTTGATTCAGA 1080
1166 GCATTTAAGCAGATTTTAAACATCGGACGATCCGAAAGCAGCAGAGAGAAATTC 1225
1081 GCATTTAAGCAGATTTTAAACATCGGACGATCCGAAAGCAGCAGAGAGAAATTC 1140
1226 AGAGCAAAACAAATTAAGGAAAGAGAGAGGACCAAGTTGGTTAAATTTATTCAGAGGA 1285
1141 AGAGCAAAACAAATTAAGGAAAGAGAGAGGACCAAGTTGGTTAAATTTATTCAGAGGA 1200
1286 AATCAAGATTTGTTAGATTAATTCATGATGATGATGATGATGATGATGATGATGATGATG 1345
1201 AATCAAGATTTGTTAGATTAATTCATGATGATGATGATGATGATGATGATGATGATGATG 1260
1346 CACCCAGATTAACCAAAACCACTTAGATTTCTTGAAGGAAATTAATGATGATGATGATGATG 1405
1261 CACCCAGATTAACCAAAACCACTTAGATTTCTTGAAGGAAATTAATGATGATGATGATGATG 1320

1406 GAGTTTGATCTGAGTCTTAACATCAATGGAGTGGTCAAAAGCTTACAAAGAAAGCCGAGTA 1465
1321 GAGTTTGATCTGAGTCTTAACATCAATGGAGTGGTCAAAAGCTTACAAAGAAAGCCGAGTA 1380
1466 GCAAACCTTCACTTTCCAAAGTGTATGTATGTAGAACAGAAAAACCAACCAATGAGACGATT 1525
1381 GCAAACCTTCACTTTCCAAAGTGTATGTATGTAGAACAGAAAAACCAACCAATGAGACGATT 1440
1526 TCTACTCTAAATCTTTTACCATCAA 1549
1441 TCTACTCTAAATCTTTTACCATCAA 1464

RESULT 7
AX098228 LOCUS 6233 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 3 from Patent WO0118208.
ACCESSION AX098228
VERSION AX098228.1 GI:13515357
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
TITLE Interferon induced polynucleotides and proteins encoded thereby
JOURNAL Patent: WO 0118208-A 3 15-MAR-2001;
Curagen Corporation (US); BIOGEN, INC. (US)

FEATURES
source
1..6233
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1..4770
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC35438.1"
/db_xref="GI:13515358"
/db_xref="REMBL:CA35438"
/translation="MAKQLNLPENTDDWTKEDVNWLSHKKIDOKREILTEQDVNGA
VLKLLKELHVDMGITHGPAIQIEELFKELRTAIEDSIQTSKMGKPSKNAPKQDTVS
QKERETSKQKQKNGENPDMAIPSTAKGSKSLVELIEDIKDYTKRQPSIDLT
CVSYPDFDSFNYKLPDLSQPGNLDIPHEFKATNTATATEDVCKPESNE
VFRFASACNSNTNGTIFHGKDPKHGKIVGKVTNDKEALINFINMINKYIFPDHQ
VOAKKCPERPEFVELLPNSLSDRFVIEDVIIPQSECOVDYFQIKMKNKNIWE
QSKPSLFRDGTSSKDTKMKVDPAKADPKTLAEKKAEEFRKNTKKREGP
KLKLLTGQDLDNSYIEQYILVTKCHPDQTKHLDLKEIKWFAVLEFDPESNIG
VYAKESKRVANLHFPVVEQKTPNETISTNLHOFPSVIFCNGLRDLDSKXKPP
DPSQWREASDVRLIISPLTHEDIMPRGKFLVFLLSVDDPDPLIETFCATQD
LKGEMNLICIVHPHIFQWKDLLEARTKHQDEISSQCISALSLEENGLIKLSV
TQSKRLPSIGLSTVLLKKEEDIMTALEIICENECEGTLLKDKNFKLEPKSEED
FYRGKVSWNFYFSSSYSPFVKRDKYERLEAMIQNCADSSKPTSTKIHLYHHPG
CGGTLAMHILMELKKERCAVLKNTVDPSFGEQVTSITYGMNROVYVPLLV
DPEQDQNVLLQYSIQTAIKYIRYKPLVILNCWPSONPEKSARIPDSIAIQQ
LSPKQAPLKEIKELKEQKPFDFYFWMKNTFNKEYIENVRNLLKGNITKE
AKLFFLALLNSYVPDITISQCEKFLGKNGKAPGTEKEDKMGITSTILLIETV
IBGNYCGVRIIHSIAEFLSEELKSKYHLNQLMQLTENLFPDGMGSKFLQD
MHTLLTRRDEHEGTGTFEALHKGDENEAVEALLLESIRHFNBNAGNISI
ARHFTKKDKFGNALNWAQAKIIPDINSYISDTLGVYKSKIRWIBENGNGNISI
DLIALLDLAEASAFKESQCSQSDREVEKERYPKSKRYDVTNIAGYQGEIEVG
LVYTLQLIPLFFDNKNSKRYMNFVSGSDIPGDNNVEYKALKNYIPYLTKLF
SUKESEPFDEYFVLLKPRNNIKQNEEAKTRKRVAGYFKYVDIFCLLESQNTGLG
SKFSEPLQVERCNRNLVALKADFGSLLELYLKQSDALSTWKCLVNETYFLBQCTV
KIQSEKLMFILANITLISCTISRLVPEKLDQREVLPQIAYPFLGKGRLERLVH
LLFPWENQDLQSEOMKQYALANKSPKGYKMHRTKQIAYPFLGKGRLERLVH
KGIKIDCPKPTPDINSLWQSGDWKEEKVQELLRLQGRAENNNLYIEYGINEKITIP
IIPAFLGQLRSGRSIEKVSFYLGSIGGLAYDIEIV"

ORIGIN

Query Match 94.4%; Score 1462.4; DB 6; Length 6233;
Best Local Similarity 99.9%; Pred. No. 3.3e-266;

Matches	1463;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	86	ATGCGAAAGCAACTTAACTTCCAGAAAAATACAGATGATTGGACAAAGAGGATGTAAT	145						
Db	1	ATGCGAAGCAACTTAACTTCCAGAAAAATACAGATGATTGGACAAAGAGGATGTAAT	60						
Qy	146	CAGTGGTTTGAAGTCTAAGATTGACCAAAAAACACAGGGAATTTTGACTGAAACAAGAC	205						
Db	61	CAGTGGTTGAAGTCTAAGATTGACCAAAAAACACAGGGAATTTTGACTGAAACAAGAC	120						
Qy	206	GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA	265						
Db	121	GTGAATGGAGCAGTCTTGAAGTGGTTAAAAAAGAACATCTTGTTGATATGGGCATCACA	180						
Qy	266	CATGGACCCAGCTATTCAAATAGAAAGACTATTCAAAGAAATTTGGGGAAAAACAGCCATTGAA	325						
Db	181	CATGGACCCAGCTATTCAAATAGAAAGACTATTCAAAGAAATTTGGGGAAAAACAGCCATTGAA	240						
Qy	326	GATTTCGATTACAGACATCTAAGATGGGAAGGCCAGTAAAAAATGCTCTTAAGACCAAACT	385						
Db	241	GATTTCGATTACAGACATCTAAGATGGGAAGGCCAGTAAAAAATGCTCTTAAGACCAAACT	300						
Qy	386	GTGTCTCAAAGGAAACGTAGAGAACTTCAAAGCAAAAAAAGGGTAAAGAGAACCCCA	445						
Db	301	GTGTCTCAAAGGAAACGTAGAGAACTTCAAAGCAAAAAAAGGGTAAAGAGAACCCCA	360						
Qy	446	GATATGCTTAATCCGTCTGCNATGAGTACAACCTGCTAAAGGTTCTAAGTCACTAAAGTT	505						
Db	361	GATATGCTTAATCCGTCTGCNATGAGTACAACCTGCTAAAGGTTCTAAGTCACTAAAGTT	420						
Qy	506	GAGTCTATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTTGACA	565						
Db	421	GAGTCTATAGAAGATAAAATAGATTATACAAAGGAAAGGCAACCATCCATAGACCTTGACA	480						
Qy	566	TGTTGATCATATCATTTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTGT	625						
Db	481	TGTTGATCATATCATTTTGATGAATTCAGTAATCCATATCGTTACAAGTTGGATTTTGT	540						
Qy	626	CTACAGCCTGAAACAGGACACAGGCAATCTCATTTGATCCGATACATGAATTTCAAAGCCTTC	685						
Db	541	CTACAGCCTGAAACAGGACACAGGCAATCTCATTTGATCCGATACATGAATTTCAAAGCCTTC	600						
Qy	686	ACAAATACAGCAACAGCCACAGAGAGAGGTGCAAGATGAAATTTACGAATGAGGTTTC	745						
Db	601	ACAAATACAGCAACAGCCACAGAGAGAGGTGCAAGATGAAATTTACGAATGAGGTTTC	660						
Qy	746	CGATTTGCTTCAGCTTGTATGAAATTCAGTACCAATGGCACTATTCATTTTGGAGTCAAA	805						
Db	661	CGATTTGCTTCAGCTTGTATGAAATTCAGTACCAATGGCACTATTCATTTTGGAGTCAAA	720						
Qy	806	GACAAACCCATGGGAAATTTTGGCATCAAGTCACCATGATGATCAAGGAGCCCTC	865						
Db	721	GACAAACCCATGGGAAATTTTGGCATCAAGTCACCATGATGATCAAGGAGCCCTC	780						
Qy	866	ATTAACCATTTCAATCTGATGATAAACAAAGTATTTTGAAGACCATCAAGTCCACAAAGCA	925						
Db	781	ATTAACCATTTCAATCTGATGATAAACAAAGTATTTTGAAGACCATCAAGTCCACAAAGCA	840						
Qy	926	AAGAGTGCATTCGAGACCCAGATTTGTGGAAGTTTACTGCGCAATATGTTACTCTATCT	985						
Db	841	AAGAGTGCATTCGAGACCCAGATTTGTGGAAGTTTACTGCGCAATATGTTACTCTATCT	900						
Qy	986	GACGATTTGTTATTGAAGTGGACATTTATCCACAGTTTCTCTGAATGCCAATATGATTAC	1045						
Db	901	GACGATTTGTTATTGAAGTGGACATTTATCCACAGTTTCTCTGAATGCCAATATGATTAC	960						
Qy	1046	TTCAGATTAATAATCCAAATTTACACAAACAAATATGGGAACCAAGTAAAAAATCTCTCA	1105						
Db	961	TTCAGATTAATAATCCAAATTTACACAAACAAATATGGGAACCAAGTAAAAAATCTCTCA	1020						
Qy	1106	CTATTGTGCGAGATGGGACAGCTCTTAGGACATTAACGAAAAATAAGTTGATTTCAGA	1165						
Db	1021	CTATTGTGCGAGATGGGACAGCTCTTAGGACATTAACGAAAAATAAGTTGATTTCAGA	1080						

Qy	1166	GCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAAAATTC	1222
Db	1081	GCATTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGCAGCAGAGAAAATTC	1149
Qy	1226	AGAGCAAAAACAAATAAAAAGAAAGAGAGGGGACCAAAGTTGGTTAAATTTATTGACAGGA	1285
Db	1141	AGAGCAAAAACAAATAAAAAGAAAGAGAGGGGACCAAAGTTGGTTAAATTTATTGACAGGA	1200
Qy	1286	AATCAAGATTTGTTAGATTAATTCATCTACTATGACAGTTCACATCTCTGTAAACAAATAAATGC	1345
Db	1201	AATCAAGATTTGTTAGATTAATTCATCTACTATGACAGTTCACATCTCTGTAAACAAATAAATGC	1260
Qy	1346	CACCCAGATCAAAACAAACAACTTTAGATTTCTCTGAAGGAAATTAATTTGGTTTGGTGTATTG	1405
Db	1261	CACCCAGATCAAAACAAACAACTTTAGATTTCTCTGAAGGAAATTAATTTGGTTTGGTGTATTG	1320
Qy	1406	GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTTCAGAGCTTACAAAGAAAGCCCGAGTA	1465
Db	1321	GAGTTTGATCCTGAGTCTAACATCAATGAGTGGTTCAGAGCTTACAAAGAAAGCCCGAGTA	1380
Qy	1466	GCAACCTTCACATTTCCAACTGTATATGTAGACAGAAAACCAACCAATGAGACGATT	1525
Db	1381	GCAACCTTCACATTTCCAACTGTATATGTAGACAGAAAACCAACCAATGAGACGATT	1440
Qy	1526	TCTACTCTAAATCTTTTACCATCAA	1549
Db	1441	TCTACTCTAAATCTTTTACCATCAA	1464
RESULT 8			
AC092513/c			
LOCUS	AC092513	176350 bp	DNA linear HTG 19-JUN-2002
DEFINITION	Papio anubis clone RP41-474E21, WORKING DRAFT SEQUENCE.		
ACCESSION	AC092513		
VERSION	AC092513.2 GI:21465392		
KEYWORDS	HTG; HTGS PHASE2; HTGS DRAFT.		
SOURCE	Papio anubis (olive baboon)		
ORGANISM	Papio anubis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopithechinae; Papio.		
AUTHORS	1 (bases 1 to 176350)		
TITLE	Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Ian, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D., McColley, J.C., McDowell, J., Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Schuler, M.G., Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
JOURNAL	NISC Comparative Sequencing Initiative		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 176350)		
TITLE	Green, E.D.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717		
AUTHORS	Government Circle, Gaithersburg, MD 20877, USA		
TITLE	3 (bases 1 to 176350)		
JOURNAL	Green, E.D.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (19-JUN-2002) NIH Intramural Sequencing Center, 8717		
TITLE	Government Circle, Gaithersburg, MD 20877, USA		
JOURNAL	On Jun 19, 2002 this sequence version replaced gi:1471326.		
COMMENT	----- Genome Center Center: NIH Intramural Sequencing Center Center code: NISC Web site: http://www.nisc.nih.gov Contact: nisc.zoo@nih.gov ----- Project Information Center project name: aml		

Center clone name: 474E21

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 176118 bases at least Q40
 Consensus quality: 176233 bases at least Q30
 Consensus quality: 176319 bases at least Q20
 Insert size: 135000; agarose-fp
 Insert size: 172000; pulse-field-gel
 Insert size: 176350; sum-of-contigs
 Quality coverage: 17.59x in Q20 bases; agarose-fp
 Quality coverage: 13.81x in Q20 bases; pulse-field-gel
 Quality coverage: 13.47x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 *
 * 1 176350: contig of 176350 bp in length.

FEATURES

```

source
  1. .176350
    /organism="Papio anubis"
    /mol_type="genomic DNA"
    /db_xref="taxon:9555"
    /clone="RP41-474E21"
    /clone_lib="RP41"

misc_feature
  1. .176350
    /notes="assembly_fragment"
    clone end:SP6
    vector side:left
    clone end:T7
    vector side:right

misc_feature
  1. .44288
    /notes="clone overlaps with GenBank Accession Number
    AC092407 clone RP41-402M16 (center project name amk)"
    136968..176350

misc_feature
    /notes="clone overlaps with GenBank Accession Number
    AC092515 clone RP41-6601 (center project name ami)"

ORIGIN
Query Match          90.2%; Score 1396.8; DB 2; Length 176350;
Best Local Similarity 96.5%; Pred. No. 4e-254;
Matches 1428; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy  70  ACTACTGGGTATCAGAAATGCGAAAGCACTTAACCTTCCAGAAATATACAGATGATGGAC 129
Db  99493 ACATGTAGGTATCAGAAATGCGAAAGCACTGAACCTTCCAGAAATATACAGATGATGGAC 99434
Qy  130  AAAAGAGGATGTAATACAGTGGTTAGAAAGTCTAAGATGACCAAAAACACAGGGAAT 189
Db  99433 CAAAGAGGATGTAATACAGTGGTTAGAAAGTCTAAGATGACCAAAAACACAGGGAAT 99374
Qy  190  TTGTGCTGACACAGACGTGAATGGACAGCTTCTGAAGTGGTTAAAAAAGAACATCTTCT 249

```

```

Db  99373 TTGTGCTGACACAGACGTGAATGGACAGCTCTTGAAGTGGTTAAAAAAGAACATCTTCT 99314
Qy  250  TGAATATGGGCATCACACATGGACCAAGCTATTTCAATAGAGAACTATTTCAAGAAATTCG 309
Db  99313 TGAATATGGGCATCACACATGGACCAAGCTATTTCAATAGAGAACTATTTCAAGAAATTCG 99254
Qy  310  GAAACACAGCCATTTGAAGATTCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAAAATGC 369
Db  99253 GAAACACAGCCATTTGAAGATTCGATTCAGACATCTTAAGATGGGAAAGCCAGTAAAAATGC 99194
Qy  370  TCCTAAAGACCAAACTGTCTCTCAAAAGGAAAGTAGAGAACTTCAAGACCAAAAACAAA 429
Db  99193 TCCTAAAGACCAAACTGTCTCTCAAAAGGAAAGTAGAGAACTTCAAGACCAAAAACAAA 99134
Qy  430  GGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCTGCAATGAGTAGCAACTGCTGTAAGGTTTC 489
Db  99133 GGGTAAAGAGAACCCAGATATGGCTTAATCCGCTCTGCAATGAGTAGCAACTGCTGTAAGGTTTC 99074
Qy  490  TAAAGTCACTAAAGTTGAGCTCATAGAGATAAATAGATTTATACAAAGGAAGGCAACC 549
Db  99073 TAAAGTCACTAAAGTTGAGCTCATAGAGATAAATAGATTTATACAAAGGAAGGCAACC 99014
Qy  550  ATCCATAGACCTGCACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGTTA 609
Db  99013 ATCCATAGACCTGCACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGTTA 98954
Qy  610  CAAAGTTGGATTTTGTAGTCTACAGCCTGAAACAGACACAGGCAATCTCATTTGATCCGATACA 669
Db  98953 CAAAGTTGGATTTTGTAGTCTACAGCCTGAAACAGACACAGGCAATCTCATTTGATCCGATACA 98894
Qy  670  TGAATTCAAAGCCTTCACAAATACAGACACAGGCAATCTCATTTGATCCGATACAAT 729
Db  98893 TGAATTCAAAGCCTTCACAAATACAGACACAGGCAATCTCATTTGATCCGATACAAT 98834
Qy  730  TAGCAATGAGGTTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTT 789
Db  98833 TAGCAATGAGGTTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTTCCGATTT 98774
Qy  790  TCATTTTGGAGTCAAAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCCACCAATGA 849
Db  98773 TCATTTTGGAGTCAAAAGACAAACCCCATGGGAAATTTGTTGGCATCAAAGTCCACCAATGA 98714
Qy  850  TACCAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCA 909
Db  98713 TACCAGGAAGCCCTCATTAACCATTTCAATCTGATGATAAACAAGTATTTTGAAGACCA 98654
Qy  910  TCAAGTCCAAACAGCAAGAAAGTGCATTCGAGAGCCAGAGTTTGTGGAAGTTTACTGCC 969
Db  98653 TCAAGTCCAAACAGCAAGAAAGTGCATTCGAGAGCCAGAGTTTGTGGAAGTTTACTGCC 98594
Qy  970  AAATAGTACTCTATCTGACAGATTTGTTTATTGAAGTGGACATTTATCCAGTTCCTCTGA 1029
Db  98593 AAATAGTACTCTATCTGACAGATTTGTTTATTGAAGTGGACATTTATCCAGTTCCTCTGA 98534
Qy  1030  ATGCCAATATGATTTATCTCCAGATTAATAATGCAAAATTAACAACAACAATATGGGAACA 1089
Db  98533 ATGCCAATATGATTTATCTCCAGATTAATAATGCAAAATTAACAACAACAATATGGGAACA 98474
Qy  1090  AAGTAAAAAATTTCTCACTATTTGTGCGAGATGGGACCCAGCTCTTAAGGACATTTACGAAAA 1149
Db  98473 AAGTAAAAAATTTCTCACTATTTGTGCGAGATGGGACCCAGCTCTTAAGGACATTTATGAAAA 98414
Qy  1150  TAAAGTTGATTTTCAGACATTTAAGCAGATTTTAAACAACCTGGCAGAGTCCAGAAAGC 1209
Db  98413 TAAAGTTGATTTTCAGACATTTAAGCAGATTTTAAACAACCTGGCAGAGTCCAGAAAGC 98354
Qy  1210  AGCAGAAAGAAAAATTCAGAGCAAAAACAAATAAAAAAGAAAGAGAGGACCAAGTTGGT 1269
Db  98353 AGCAGAAAGAAAAATTCAGAGCAAAAACAAATAAAAAAGAAAGAGGACCAAGTTGGT 98294
Qy  1270  TAAATTTATGACAGAAATCAAGATTTGTTAGATTAATTTATCTATGAAACAGTACATCT 1329
Db  98293 TAAATTTATGACAGAAATCAAGATTTGTTAGATTAATTTATCTATGAAACAGTACATCT 98234

```


QY 257 GGCATCACATGACGACGCTATTCAATAGAGAACTATTCAAGAAATTCGCGAACA 316
Db 170404 GGTATAACATTTGGGCCAGCTATCCAGATAGAACTATTATTCAAGAGTTGATGGAACA 170345
QY 317 OCCATTGAAGATTGATTCAGACATCTAAGATGGGAAAGCCAGTAAATAATGCTCTTAAA 376
Db 170344 TCCACTGAAGATCTTCTCAGATTGTAAGAGGGGAAAGCCAGTAAAGATGTTCTCTAAA 170285
QY 377 GACCAACTGTGCTCAAAAGGACGTAGAGAACTTCAAGAGCAAAAACAAAGGGTAAA 436
Db 170284 AA-----ACAAGAGAAAACAGAGAACTTCAAGAGCAAAAACAAAGGATAAA 170237
QY 437 GAGAACCCAGATATGGCTTAATCCGCTGCAATGAGTCAACCTGCTAAAGGTTCTTAAGTCA 496
Db 170236 GAGACATCAGATAGAGTTACTGACTCTACATATAGTACAGTTACTGAAGTTCTATGCA 170177
QY 497 CTAAAGTTGAGCTCATAGAGATAAATAGATTATACAAAGGAAGCAACCATCAATA 556
Db 170176 CTTAACAATGAATTCATGGAAGTGAATAAATGACACACAGAAAAAGCAGTCACTGTGA 170117
QY 557 GACCTGACATGTATATATCCATTTGATGAATTCAGTATCCATATCGTTACAAGTTG 616
Db 170116 GAACATGATGTGTACCATACCTTTTGTATGAATTCAGTATCCATATCGTTACAATG 170057
QY 617 GATTTAGTCTACAGCTTGAACACAGGACCGCAATCTCATTTGATCCGATACATGAATTC 676
Db 170056 AATTTAGTCTACAGCTTGAACACAGGACCACTCAATCTCATTTGATCCATATCAATG 169997
QY 677 AAAGCTTCACAAATACAGACAGCCACAGAGGAGTGTCAAGATGAATTTAGCAAT 736
Db 169996 AAAGCTTCACAAATACAGACAGCCACAGAGGAGTGTAGATGAATTTAGCAAT 169937
QY 737 GAGGTTTTCCGATTTGCTTACGCTTGTATGAATTCACGTACCAATGCGCACTATTCTTT 796
Db 169936 GAGGTTTTCCGATTTGCTTACGCTTGTATGAATTCACGTACCAATGCGCACTATTCTTT 169877
QY 797 GGAGTCAAGACAAACCCCATGGGAAATTTGTGGCATCAAGTCCCAATGATACCAAG 856
Db 169876 GGAGTCAAGACAAACCCCATGGGAAATTTGTGGCGGGAAGTCCACACTGTCCACCAAG 169817
QY 857 GAAGCCCTCATTAACCATTTCAATCTGATGATAACAGATATTTGAAGACCATCAAGTC 916
Db 169816 GAAGCCCTCATTTGACCATTCATATATGATGATCCACAGTATTTGAAGACCATCAGTC 169757
QY 917 CAAGCAAGAAAGATGATTCGAGAGCCAGATTTGTGGAGTTTACTGCCAAATAGT 976
Db 169756 CAAAAGCGAAGAACTGCAATTCGAGAGCCAGATTTGTAGAAGTTTACTGCCAAACAGT 169697
QY 977 ACTCTATCTGACAGATTTGTTATTGAAGTGGACATTTATCCACAGTTCTCTGAATGCCAA 1036
Db 169696 ATTCTATCTGACAGATTTGTTATTGAAGTGGATGTTGTTCCAAATATTTCTGAATGTGA 169637
QY 1037 TATGATTAATCTCAGATTAATAATGAAAATTACAAACAAAATATGGGAACAAAGTAAA 1096
Db 169636 CAAGAGTATTTCCAGATTAATAATGAAAATTACAAACAAAATATGGGAACAAAGTCCA 169577
QY 1097 AAATTTCTACTATTTGTGGAGATGGGACGCTCTAGGACATTCAGAAAATAAAGTT 1156
Db 169576 AAATTTCTAGTCTTCATGCGAGATGGGCGAGTTCTTAGGACGCTGATGAAAATAAATGTA 169517
QY 1157 GATTTTCAGAGCATTTTAAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAAGCAGCAGAA 1216
Db 169516 GAATTTCAAAACATTTTAACTGGATTTAAAAAACTGGCAGAGCTAGGAAAGAGCAGAA 169457
QY 1217 GAAAATTCAGAGCAAAAACAAATAAAAAGAAAGAGGACCAAGTGTGTTAATTA 1276
Db 169456 GAAAATTCAGACTGAAAACAAACAAAAGAGAGTGGGACCAAACTGTTAAATG 169397
QY 1277 TTGACAGAGAAATCAAGATTTGTTAGATAATTCATATGATGAACAGTACATTTCTGTAA 1336
Db 169396 TTAAACAGGAAATAGGATTTGTTAGATAATTCATATGATGATGCTGTGATCAATTTCTGTA 169337

QY 1337 AATAAATGCCACCCAGATCAACAAAAACACTTAGATTTCTCGAAGAAATTTAAATGGTTT 1396
Db 169336 AATAAATGCCATCCGACTCAGATAAAAACACTTAGATTTCTTAAAGAAATTTAAGTGGTTT 169277
QY 1397 GCTGTATTGGAGTTTATCCTCGAGTCTAAACATCAATGGAGTGGTCAAGCTTTACAAAGAA 1456
Db 169276 GCTGTATTGGAGTTTATCCTCGATCTGTAGTGAAGTGTGGTCAAGCTTTACAAAGAA 169217
QY 1457 ACCCGAGTAGCAAACTTCACATTTCCAGTGTATATGTAGAACAGAAACCAACCAAT 1516
Db 169216 ACCCGAGTAGTAACTTCACATTTCCAGTGTATATGTAGAGGGGAAACCA---CAAT 169160
QY 1517 GAGACGATTTCTACTCTAAATCTTTACCATCAA 1549
Db 169159 GAGACGATTTAGTCTGAATCTTTATCAACAA 169127

RESULT 10
AC108893/c
LOCUS
DEFINITION
AC108893
AC108893.2 GI:21909465
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Ruteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 108538)
Akther, N., Antonellis, A., Ayele, K., Becketrom-Sternberg, S. M.,
Benjamin, B., Blakesley, R., Bouffard, G. G., Breen, K., Brinkley, C.,
Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J.,
Haghigbi, P., Hansen, N., Ho, S.-L., Idol, J. R., Karlins, E., Laric, P.,
Lee-Lin, S.-Q., Legaspi, R., Maduro, Q. L., Maduro, V. B.,
Marquillas, E. H., Masello, C., Maskeri, B., Mastrian, S. D.,
McCloskey, J. C., McDowell, J., McGuire, C., Pearson, R.,
Portnoy, M. E., Prasad, A., Schueler, M. G., Stantrop, S., Thomas, J. W.,
Thomas, P. J., Touchman, J. W., Tsurgon, C., Vogt, J. L., Walker, M. A.,
Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE
2 (bases 1 to 108538)
Green, E. D.
Direct Submission
Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE
3 (bases 1 to 108538)
Green, E. D.
Direct Submission
Submitted (19-JUL-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Jul 19, 2002 this sequence version replaced gi:18464069.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: CJO
Center clone name: 500608

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated

with a Phrap-derived quality score.

----- Summary Statistics -----
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 108439 bases at least Q40
 Consensus quality: 108499 bases at least Q30
 Consensus quality: 108531 bases at least Q20
 Insert size: 105000; agarose-fp
 Insert size: 108538; sum-of-ctnigs
 Quality coverage: 13.02x in Q20 bases; agarose-fp
 Quality coverage: 12.60x in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * *
 * 1 108538: contig of 108538 bp in length.

FEATURES

source

1..108538
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="RP42-500G8"
 /clone_lib="RP42"

misc_feature

1..108538
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 67.1%; Score 1039.8; DB 2; Length 108538;
 Best Local Similarity 83.2%; Pred. No. 18-186;
 Matches 1227; Conservative 0; Mismatches 227; Indels 21; Gaps 3;
 75 TGGGTATCAAGATGGCAAGCAACTTAACTTCCAGAAAATACAGATGTCGACAAAG 134
 Db TAGGTATCAAGATGGCAGCACAACCTTAACTTCCAGAAAATACAGATGTCGACAAAG 53712
 135 AGGATGTAATCGATGGTGTAGAAAGTCATAAGTTGACCAAAAACACAGGGAAATTTGA 194
 Db AGGATGTAATCGATGGTGTAGAAAGTCATAAGTTGACCAAAAACACAGAGCAATTTGA 53652
 195 CTGAACAAGACGTGAATGGAGCAGCTCTTGAAGTGGTTAAAAAAGAACATCTTTGTGATA 254
 Db CTGCAACAGATGTGAGTGGTGAATCTTGAAGTACTTAACTAAAGATCACCTTTGTGCTA 53592
 255 TGGGATACACATGACACGACGCTATTTCAAATAGAGAACTATTCAAGAAATGGCGAAA 314
 Db TGGGATACACATTTGGACACGCTATTTCAAATAGAGAACTATTTCAAAGAGTTGCTGGAGA 53532
 315 CAGCATTGAAGATTCGATTTCAGACATCTAAGATGGGAAAGCCAGTAAAAATGCTCCTA 374
 Db CATCTCTGAGATCTCTTCCAGACATGTAAAGCGGAAAGCGCAGTAAAGTGTCTCTA 53472
 375 AAGACCAACTGTGCTTCAAAAGGAACGTTAGAGAACTTCAAGCAAAAACAAAAGGGTA 434
 Db AAGACCAACTGTGCTTCAAAAGGAACGTTAGAGAACTTCAAGCAAAAACAAAAGGGTA 53430
 435 AAGAGAACCCAGATATGGCTTAATCCCTGCAATAGTATCAACTGCTGAAGGTTCTAAGT 494
 Db AAGAGAAATCAGATAGGTTAATGACTCTACAGTGAAGTACTTAAAGGTTCTAAGT 53370
 495 CACTAAAGTTGAGCTCATAGAGATAAAATAGATTATCAAGGAAAGGCAACCATCA 554
 Db CACTGAAAAATGAGTTTCAATGGAAGATGAATAGATGACACAAAGAAAAAGCAGCCATCA 53310

QY 555 TAGACCTGACATGTGTATCATATCCATTTGATGAATTCAGTAAATCCATATCGTTCAAGT 614
 Db 53309 CAGAACCGACATGCGATGCGATACCCCTTTTGATGAATTCAGTATCGTATCGTTCAAAAT 53250
 QY 615 TGGATTTTGTATGACAGCTGAAACAGGACCGAGGCAATCTCATTTGATCCGATACATGAAT 674
 Db 53249 TGAATTTTAACTTACAGCTGAAACAGGACCGACCTCAATCTCATTTGATCCAAATACATGAAT 53190
 QY 675 TCAAGCCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAGATGAAATTTAGCA 734
 Db 53189 TCAAGCCCTTCACAAATACAGCAACAGCCACAGGACGAGAGGATGCTAAATTTGAATTTAGCA 53130
 QY 735 ATGAGTTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGCGACATTTCAAT 794
 Db 53129 ATGAGTTTTCCGATTTGCTTTCAGCTTGTATGAATTCACGTACCAATGCGACATTTCAAT 53070
 QY 795 TTGGATTCAGAGCAACACCCCATGGCAAAATTTGTCGCATCAAAAGTCACCAATGATACCA 854
 Db 53069 TTGGATTCAGAGCAACACCTTCAGGAACTAATTTGTCGCTGGAAATTTACCAATCACC 53010
 QY 855 AGGAAGCCCTTCATTAACCAATTTCAATCTGATGATAAAACAAAGTATTTGAAGACCATCAAG 914
 Db 53009 AGGAAGCCCTTCATTAACCAATTTCAATCTGATGATAAAACAAATTTGAAGACCATCAGG 52950
 QY 915 TCCAAACAGCAAGAAAGTCGATTCGAGAGCCAGATTTGTGGAAGTTTCTGCAAAATA 974
 Db 52949 TCCAAACAGCAAGAAAGTCGATTCGAGAGCCAGATTTGTGGAAGTTTCTGCAAAATA 52890
 QY 975 GTACTCTATCTGACAGATTTGTTTATTTGAAGTGACATTTTCCACAGTTTCTCTGAATGCC 1034
 Db 52889 GTACTCTATCTGACAGATTTGTTTATTTGAAGTGACATTTTCCACAGTTTCTCTGAATGCC 52830
 QY 1035 AATATGATTTCTCCAGATTTAAATGCAAAATTTACCAACAAATATTTGGACCAAGTA 1094
 Db 52829 AAGTTGATTTATTTCCAGATTTAAATGCAAAATTTTCAGACACACAAATGGAACAAAGTC 52770
 QY 1095 AAAAAATCTCACTATTGTCGAGATGGGACCAAGCTCTAAGGACATTTACGAAAAATAAAG 1154
 Db 52769 CAAACTTCTCAGTCTTTGTCGAGATGGGACCAAGTTCTAAGGACATCATGAAAAATAA 52710
 QY 1155 TTGATTTTCAGACATTTTAAGCAGATTTTAAACACTGGCAGAGTCCAGAAAGACGACAG 1214
 Db 52709 TGGATTTCAAGCATTTAAATTAGATTGAAAAGACTTGCAGAACTTTCGAAAGAGCGCAG 52650
 QY 1215 AAGAAAAATTCAGAGCAAAAAACAAATAAAAAAGAGAGAGGACCAAAAGTTTGGTTAAAT 1274
 Db 52649 AAGAAAAATTCAGAGTCAAAAAACAAATAAAAAAGAGAGGAGGACCAAAAGTTTGGTTAAAT 52590
 QY 1275 TATTGACAGAAATCAAGATTTGTTAGATTAATTCATATCTATGAAACAGTACATTTCTTTGTA 1334
 Db 52589 TGTTCACAGAAATCAAGATTTGTTAGATTAATTCATATCTATGAAACAGTACATTTCTTTGTA 52530
 QY 1335 CAAATAAATGCCACCGACGATCAAAACAAACACATTTAGATTTCTGGAAGGAAATTTAAATGGT 1394
 Db 52529 TAAATAAATGCCACCGCTTACAAATCAACACATTTAGATTTCTTAAAGGAAATTTAAATGGT 52470
 QY 1395 TTGCTGTATTGGAGTTTGTATCCTGAGTCTAAACATCAATGAGAGTGGTCAAGCTTCAAG 1454
 Db 52469 TTGCTGTATTGGAGTTTGTATCCTGAGTCTAAACATCAATGAGAGTGGTCAAGCTTCAAG 52410
 QY 1455 AAAGCCAGTACGAAACCTTCACTTTCCAAAGTGTATATGTAGAAACAGAAACCAACCA 1514
 Db 52409 AAAGCCAGTACGAAACCTTCACTTTCCAAAGTGTATATGTAGAAACAGAAACCAACCA 52353
 QY 1515 ATGAGACGATTTCTTACTCTAAATCTTTTACCATCA 1549
 Db 52352 CTGAGAGATTTACTAGTCTGAATCTTTTATCAACAA 52318

RESULT 11
 AC079885/c
 LOCUS

AC079885 148278 bp DNA linear HTG 11-SEP-2002

```

DEFINITION      Rattus norvegicus chromosome 4 clone RP31-246H18 strain Brown
Norway, WORKING DRAFT SEQUENCE, 6 ordered pieces.
ACCESSION       AC079885
VERSION         AC079885.3 GI:22779511
KEYWORDS        HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE          Rattus norvegicus (Norway rat)
ORGANISM        Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 148278)
Ahter.N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaapi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,K.,
Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
Stantropop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 148278)
Green,S.D.
Direct Submission
Submitted (16-SEP-2000) NIH Intramural Sequencing Center, 8717
Grosvmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 148278)
Green,E.D.
Direct Submission
Submitted (11-SEP-2002) NIH Intramural Sequencing Center, 8717
Grosvmont Circle, Gaithersburg, MD 20877, USA
On Sep 11, 2002 this sequence version replaced gi:11120768.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoehngri.nih.gov
----- Project Information
Center project name: rk
Center clone name: 246H18

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8X average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 146350 bases at least Q40
Consensus quality: 147034 bases at least Q30
Consensus quality: 147462 bases at least Q20
Insert size: 145000; agarose-fp
Quality coverage: 11.71x in Q20 bases; agarose-fp
Quality coverage: 11.49x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 18266: contig of 18266 bp in length
18267 18366: gap of unknown length
18367 42762: contig of 24396 bp in length
42763 42862: gap of unknown length
42863 83567: contig of 40705 bp in length
83568 83667: gap of unknown length
83668 121422: contig of 37755 bp in length
121423 121522: gap of unknown length
121523 127993: contig of 6471 bp in length
127994 128094: gap of unknown length
128094 148278: contig of 20185 bp in length.

Location/Qualifiers
1..148278
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="Brown Norway"
/db_xref="taxon:10116"
/chromosome="4"
/clone_lib="RP31"
/clone_lib="RP31"
1..18266
/note="assembly_fragment
clone_end:SP6
vector_side:left"
1..3899
/note="clone overlaps with GenBank Accession Number
AC079437 clone RP31-275K4 (center project name rj)"
18367..42762
/note="assembly_fragment"
42863..83567
/note="assembly_fragment"
78240..148278
/note="clone overlaps with GenBank Accession Number
AC084240 clone RP31-103L21 (center project name rl)"
83668..121422
/note="assembly_fragment"
121523..127993
/note="assembly_fragment"
128094..148278
/note="assembly_fragment
clone_end:T7
vector_side:right"

misc_feature 72 TACTGGGTATCAGATGGCAAGCAACTTAACTTCCAGAAATACAGATGATTGACAA 131
114323 TACAGCTAATCAGATGGCGAGACACTCACTTCCAAAACACAGATGACTGGACAA 114264
132 AAGAGGATGTAATCAGTGGTTAGAAAGTCATPAAGATTGACCAAAACACAGGAAATTT 191
114263 AAGAGGATGTAATCAGTGGTTAGAAAGTCATPAAGATTGACCAAAACACAGGAAATCT 114204
192 TGACTGAACAAGACGTGATGAGCAGCTTGAAGTGGTTAAATAAAGACACATCTTGTTG 251
114203 TGATGGCAACAAGATGAGTGGAGTAGTCTTGAAGTGGTTAAATAAATAAATCTTGTTG 114144
252 ATATGGGCATCACATGACGACGCTATTCAATATAGAAGAACTATTCAAGAAGATTGCGGA 311
114143 AATATGGCATATACCCATGGCCAGCTATCCAAATAGAGAGCTATTCAAGAAGATTGCAGA 114084
312 AAACAGCCATTGAGATTTCGATTTCAGACATCTTAAGATGGGAAGCCAGTAAATGCTC 371
114083 AAACACCCCAAGAAAAATCTTACTAAGACATGTAAAGACCAAAAGCGCAAAAATATCC 114024
372 CTAAAGACCAAACTGTGTCTCAAAAGGAACTGTAGAGAACTTCAAGACCAAAAGCG 431

ORIGIN
Query Match 59.5%; Score 921; DB 2; Length 148278;
Best Local Similarity 77.6%; Pred. No. 2.6e-164;
Matches 1173; Conservative 0; Mismatches 290; Indels 48; Gaps 3;

Db 114023 CTAAGACCCAGAC-----AGAAAGTAGAGAGACTTCAAAATCAAAACCCAGAGAG 113976
Qy 432 GTAAAGAGAACCCAGATATGCTTAATCCGCTCTCAATG----- 469
Db 113975 CTGAAGAGAACTCATGCAAGGCTGACACCTCTCAGTGAGTCAGAAAGAACTCACAA 113916
Qy 470 -----AGTACNACTGCTAAAGGTTCTAAGTCACTAAAGTTGAGCTCATAGAG 518
Db 113915 GGAAGAGAACTCATGCAAGGCTAAAGGTTCTAATCAGAAAAAATGACCTAGTAGAAG 113856
Qy 519 ATAAATATAGATTATACAAAGGAAGCAACATCCATAGACCTGACATGTGTATCATATC 578
Db 113855 ATAAAGCAGTCCA--ACAGATAGAGCAGGCACTCTCCAGACCAACGTCATGATATC 113799
Qy 579 CATTTGATGATTCAGTAATCCATTCGTTTACAGTTGGATTTTGTCTACAGCTTGAAA 638
Db 113798 CTTTCAATGATTCAGTAATCCATTCGTTTACAAATTTACATTTTCCTACAGCTTGAAA 113739
Qy 639 CAGACACGAGCAATCTCATTTGATCCGATACATGAATTTCAAGGCTTTCACAAATACAGCAA 698
Db 113738 CAGGCGCACTCAATCTCATTTGATCCGATACATGAATTTCAAGGCTTTCACAAATACAGCAA 113679
Qy 699 CAGCCACAGAGAGATGTCAGATGAATTTAGCAATGAGTTTCCGATTTGCTTCAG 758
Db 113678 CAGCCACAGAGAGATGTCAGATGAATTTAGCAATGAGTTTCCGATTTGCTTCAG 113619
Qy 759 CTTGTATGAATTCACGTACCAATGGCACTATTCTATTTGGAGTCAAGACAAACCCCATG 818
Db 113618 CTTGTATGAATTCACGTACCAATGGCACTATTCTATTTGGAGTCAAGACAAACCCCATG 113559
Qy 819 GGAATTTGTTGGATCAAAAGTCACCAATGATACCAAGAGCCCTCATTAACCATTTCA 878
Db 113558 GAAGGATTTGTTGGATCAAAAGTCACCAATGATACCAAGAGCCCTCATTAACCATTTCA 113499
Qy 879 ATCTGATGATAAACAAGTATTTGAGACCATCAAGTCCAAAGCAAGAAAGAGTGCATTC 938
Db 113498 ATCTGATGATAAACAAGTATTTGAGACCATCAAGTCCAAAGCAAGAAAGAGTGCATTC 113439
Qy 939 GAGACCAAGATTTGGAAGTTTACTGCAATAGTACTPATCTGACAGATTGTGA 998
Db 113438 GAGACCAAGATTTGGAAGTTTACTGCAATAGTACTPATCTGACAGATTGTGA 113379
Qy 999 TTGAAGTGGACATTTATCCACAGTCTCTGCAATGCAATATGATTTCTCCAGATTAAA 1058
Db 113378 TTGAAGTGGACATTTATCCACAGTCTCTGCAATGCAATATGATTTCTCCAGATTAAA 113319
Qy 1059 TGCAAAATTACAAACAAATATGGGAACAAAGTAAATAATTTCTCATTTTGTGCGAG 1118
Db 113318 TGCAATGGCATAAGGATGAAACATGGCAAAATTCAAAATACTCGGTCTTTGTGAGAG 113259
Qy 1119 ATGGACACGCTTAAGGACATTAAGCAATTAAGTTGATTTTCAGAGCATTTAAGCAG 1178
Db 113258 ATGGTCCGAAAGTAAGGACATCATAGGGAATTAAGTCTGATTTCAAAGCTTTTAAGTTAG 113199
Qy 1179 ATTTTAAACACCTGGCAGAGTCAGAAAGCAGCAGAGAAATAATTCAGAGCAAAACAA 1238
Db 113198 ACCTTAAAGCAGCTGGCAGAGTCAGAAAGCAGCAGAGAAATAATTCAGAGCAAAACAA 113139
Qy 1239 ATAAAGAAAGAGAGGACCAAGTTGGTTAAATTTATTCAGAGAAATCAAGATTGTT 1298
Db 113138 ATGAAGCAACAGTCAGGGGCAAAAGCTAATTCACCTGTTGACAGAAATAAGGATCTGC 113079
Qy 1299 TAGATAATTCATATGAAACAGTACATTTCTTGAACAAATAATGCCACCCAGATCAAA 1358
Db 113078 TAGATAATTCATATGAAACAGTACATTTCTTGAACAAATAATGCCACCCAGATCAAA 113019
Qy 1359 CAAACACTTAGATTTCTGAGAGAAATTAATTTGGTTGCTGATTTGAGGTTGATCTG 1418
Db 113018 CCAACACCTTAGATTTCTGAGAGAAATTAATTTGGTTGCTGATTTGAGGTTGATCTG 112959
Qy 1419 AGTCTAATCAATGAGTGGTCAAGAGCTTACAAAGAAAGCCAGTAGCAACCTTCACT 1478
Db 112958 AGTCTGAGCAATGTTGTTGTTCAAGGCTTCAAGGAGCCAGTAGCAACCTTCACT 112899

Qy 1479 TTCCAAGTGTATGTAGACAGAAAAACCCACCAATGAGAGATTTCCTACTTAATC 1538
Db 112898 TTCCAAGCTATTATTATAGAGAAAAACCCACAGTAGTAAGTGAAGATTTCAGACTGAATC 112839
Qy 1539 TTTTACCATCAA 1549
Db 112838 TTTATCAGCAA 112828

RESULT 12
AC084240/c

LOCUS AC084240 180665 bp DNA linear HTG 06-DEC-2002
DEFINITION Rattus norvegicus chromosome 4 clone RP31-103L21 strain Brown
Norway, WORKING DRAFT SEQUENCE, 4 ordered pieces.

ACCESSION AC084240.2 GI:26080471

VERSION HTG; HTGS PHASE2; HTGS DRAFT.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 180665)

AUTHORS

Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Cariaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
Hachiguchi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Maguiles, E.H., Masello, C., Maskeri, B., McDowell, J.,
Paquitrigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,
Standripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 180665)

Green, E.D.

Direct Submission

Submitted (18-OCT-2000) NIH Intramural Sequencing Center, 8717

Grovenmont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 180665)

Green, E.D.

Direct Submission

Submitted (06-DEC-2002) NIH Intramural Sequencing Center, 8717

Grovenmont Circle, Gaithersburg, MD 20877, USA

On Dec 6, 2002 this sequence version replaced gi:10864175.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc.zoo@nigri.nih.gov

----- Project Information

Center project name: rl

Center clone name: 103L21

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 179812 bases at least Q40

Consensus quality: 180152 bases at least Q30
 Consensus quality: 180275 bases at least Q20
 Insert size: 138000; agarose-fp
 Insert size: 160000; pulse-field-gel
 Insert size: 180365; sum-of-contigs
 Quality coverage: 15.11x in Q20 bases; agarose-fp
 Quality coverage: 13.03x in Q20 bases; pulse-field-gel
 Quality coverage: 11.56x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 5391: contig of 5391 bp in length
 * 5392 5491: gap of unknown length
 * 5492 51132: contig of 45641 bp in length
 * 51133 51232: gap of unknown length
 * 51233 146593: contig of 93361 bp in length
 * 146594 146593: gap of unknown length
 * 146694 180665: contig of 33972 bp in length.

FEATURES

source

Location/Qualifiers

1..180665
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="Brown Norway"
 /db_xref="taxon:10116"
 /chromosome="4"
 /clone="RP31-103121"
 /clone_lib="RP31"
 1..71303
 /note="clone overlaps with GenBank Accession Number
 AC079885 clone RP31-246H18 (center project name rk)"
 1..5391
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 5492..51132
 /note="assembly_fragment"
 51233..146593
 /note="assembly_fragment"
 146694..180665
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 59.5%; Score 921; DB 2; Length 180665;
 Best Local Similarity 77.6%; Pred.No. 2.5e-164;
 Matches 1173; Conservative 0; Mismatches 290; Indels 48; Gaps 3;
 QY 72 TACTGGGTATCAGATGGCAAGCACTTAACTTCCAGAAAATACAGATGATTGGCAA 131
 DB 37435 TACAGCTATCAGATGGCAGAGACACTCAACTTCCAAAACACAGATGACTGGCAA 37376
 QY 132 AGAGAGATGAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTT 191
 DB 37375 AAGAGATGAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAAACACAGGGAATTT 37316
 QY 192 TGACTCAACAGAGCTGAATGGAGCAGCTCTTGAAGTGGTTAAAAAAGAACATCTTTGTTG 251
 DB 37315 TGAATGCAACAGATGAGTGGGAGTAGTCTTGAAGTGGTTAAATAAAAAATCTTTGTTG 37256
 QY 252 ATATGGGATCACAATGACAGCAGCTATTCAAAATAGAGAACTATTCAAAGAAATTCGCGA 311
 DB 37255 AATGGGCATACCCATGGGCCAGCTATCCAAATAGAGAGCTATTCAAAGAAATTCGAGA 37196
 QY 312 AAACGCCATTGAAGATTCGATTTCAGACATCTAAGATGGGAAGCCCAAGTAAATGCTC 371

DB 37195 AAACACCCCAAGAAAATCTTACTAAGACATGTAAAGACAAAAGGCAGAAAAATATCC 37136
 QY 372 CTAAAGACCAAACTGTGTCTCAAAAGGAACTGTAGAGAAACTTCAAGCAAAAACAAAGC 431
 DB 37135 CTAAAAACCCAGAC-----AGAAAGTAGAGAGACTTTCAAAATCAAAACCCAGAGAG 37088
 QY 432 GTAAAGAGAAACCCAGATATGGCTAATCCGTCTGCAATG----- 469
 DB 37087 CTGAAGAGAACTCATGCAAGGCTGACACTCTGCAAGTCCAGAGGAACTCACAA 37028
 QY 470 -----AGTACAACCTGCTAAAGGTTCTTAAGTCACTAAAGTTGAGCTCATAGAAG 518
 DB 37027 GGAAGTCCAGCAAGTCCAGAGGCTTAAAGGTTCTTAAATTCAGAAAAAATGACCTTAGTAGAG 36968
 QY 519 ATAAATAGATTATACAAAGGAAAGCGACCACTCCATAGACCTGACATGTGTATCATATC 578
 DB 36967 ATAAAGCACTCA---ACAGGATAGCAGGCACTCTCCAGGACCAAGTGCATATGCATATC 36911
 QY 579 CATTTGATGAATTCAGTAAATCATATCGTTACGTTAACAAGTTGAGTTTGTAGTTTACAGCCTGAAA 638
 DB 36910 CTTTCAATGAGTTCAAGTAAATCCATATCGTTTACAAATTTACATTTCTCTACAGCCTGAAA 36851
 QY 639 CAGGACCAAGCAATCTCATTGATCCGATACATGAATTTCAAGCCCTTCAAAAATACAGCAA 698
 DB 36850 CAGGCCATCTCAATCTCAATGACCCAGTACATGAGTTCAAGCCCTTAAACACACAGAGA 36791
 QY 699 CAGCCACAGAGAGGATGTAAGATGAATAATTTAGCAATGAGGTTTCCGATTTGCTTCAG 758
 DB 36790 CAGCCACAGAGGAGACATAAAGATGAATAATTTAGCAACGAAGTCTTCCGATTTGCTTCAG 36731
 QY 759 CTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTGGAGTCAAGAGCAAAACCCCATG 818
 DB 36730 CTTGTATGAATTCAGTACCAATGGCACTATTTCATTTTGGAGTCAAGAGCAAAACCCCATG 36671
 QY 819 GGAATAATTTGTTGGCATCAAAAGTCACCAATGATACCAAGGAAGCCCTCATTAACCAATTTCA 878
 DB 36670 GAAGGATTTGTTGGCATGAGAGTCACCAACCGTCACTAAAGATGCCCTCATTTGACCACTTCA 36611
 QY 879 ATCTGATGATAACAAGATTTTGAAGACATCAAGTCCAAAGCAAGCAAGCAAGAGTGCATTC 938
 DB 36610 ATCTGATGATAGTCCAGTATTTGAGGTCATCAAGTCCAAAGCAAGCAAGCAAGAGTGCATTC 36551
 QY 939 GAGAGCCAAAGATTTGTTGAAGTTTACTGCCAAATAGTACTCTATCTGACAGATTTGTTA 998
 DB 36550 GAGAGCCAAAGATTTGTTGAAGTTTACTGCCAAATAGTACTCTATCTAATAAATTTGTTA 36491
 QY 999 TTCAAGTGGACATTTTCCACAGTCTCTGAATGCCAATATGATTTACTTCCAGATTAATA 1058
 DB 36490 TTGAGGTGGATATCATTTCCACAAATATTCAGAAATGCAAGAGAGATTTCTTCCAAATAAAAA 36431
 QY 1059 TGCAAAATTTCAACAAACAAAATATGGGAACAAAGTAAAAAATTTCTCATTTTGTGCGAG 1118
 DB 36430 TGCAATTTGGCATAGGATGAACATGCGCAACAAATTTCCAAATACTCGGCTTTTGTGAGAG 36371
 QY 1119 ATGGGACCAAGCTTAAGGACATTTACGAAAAATTAAGTTGATTTTCAGAGCATTTAAAGCAG 1178
 DB 36370 ATGGTCCGAAAAGTAGGACATCATAGGGAATAAAGCTGATTTCAAGCTTTTAAAGTTAG 36311
 QY 1179 ATTTTAAACACATGGCAGAGTCCAGAAAAAGCAGACAGAAAAAATTCAGAGCAAAAAACAA 1238
 DB 36310 ACTTAAAGCACTGGCAGACTCTAGAAAGAGCAGAAAAAATTCAAAGTAAATCAG 36251
 QY 1239 ATAAAGAAAGAGAGGGGACCAAGTGGTTAAATTTATGACAGGAATTCAGATTTGT 1298
 DB 36250 ATGGAAGCAACAGTCAGGGGCAAAAGCTAATTTGACCTGTGACAGGAATAAGGATCTCG 36191
 QY 1299 TAGATAATTCATACTATGAACAGTAGATCTTGTAAATAAATGCCCAAGATCAAA 1358
 DB 36190 TAGATAATTCATACTATGAATGATGATCTTGTAAATAAATTCATCCAGATCAAA 36131
 QY 1359 CAAACACCTTAGATTTCTGGAAGAAATTAATATGTTTGTGTTGATTTGAGATTTGATCTCTG 1418
 DB 36130 CCAACACCTTAGATTTCTGGAAGAAATTAATATGTTTGTGTTGATTTGATCTCTG 36071

QY 1419 AGTCTACATCAATGAGTGGTCAAAAGCTTACAAAGAACCCAGTAGCAAACTTCACCT 1478
 Db 36070 AGTCTGAGACATATGTTGGTCAAGCGTTTCAAGAGGAAGCCAGTAGCAAACTTCACCT 36011
 QY 1479 TTCCAAAGTGTATGTAGAACAGAAACACACACCAATGAGAGCGATTTCTACTCTAAATC 1538
 Db 36010 TTCCAAAGCTATTATAGAGAAACACCAACAGTAGTGAAGAATTTCCAGTCTGAATC 35951
 QY 1539 TTTTACCATCAA 1549
 Db 35950 TTTTACGACAA 35940

RESULT 13
 AC134055/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-158021, *** SEQUENCING IN PROGRESS
 *** 3 unordered pieces.
 AC134055
 VERSION AC134055.3 GI:30522024
 KEYWORDS HTG; HTGS PHAS1; HTGS DRAFT; HTGS ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 221557)
 Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Bryant,N., Buhat,C., Burch,P., Burrell,K., Caldwell,E., Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Fallis,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hayes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulesed,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavjevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakolameh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pastarnak,S., Paul,H., Perez,A., Perez,L., Prankoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Fu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Stton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanik,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.
 Direct Submission
 Unpublished
 2 (bases 1 to 221557)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 221557)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:24955744.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GHAT
 Center clone name: CH230-158021
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 215241 bases at least Q40
 Consensus quality: 217114 bases at least Q30
 Consensus quality: 218453 bases at least Q20
 Estimated insert size: 248512; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 5505: contig of 5505 bp in length
 * 5506 5605: gap of unknown length
 * 5606 220233: contig of 214628 bp in length
 * 220234 220333: gap of unknown length
 * 220334 221557: contig of 1224 bp in length.
 Location/Qualifiers
 1. 221557
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-158021"
 1. 2522
 /note="wgs_end_extension
 clone_end:T7"
 232. 1034
 /note="clone_boundary
 misc_feature
 misc_feature
 source
 FEATURES
 misc_feature
 misc_feature


```
clone end:17
site:EcoRI
end_sequence: BH313763"
2573..5505
/notes="wgs_contig"
36039..62916
/notes="clone_boundary"
clone end:sp6
site:EcoRI
end_sequence: BH313765"
218385..220233
/notes="wgs_end_extension"
clone end:Sp6"

ORIGIN
Query Match 59.5%; Score 921; DB 2; Length 221557;
Best Local Similarity 77.6%; Pred No. 2.4e-164;
Matches 1173; Conservative 0; Mismatches 290; Indels 48; Gaps 3;

QY 72 TACTGGGTATCAGATGGCAAGCAACTTAACTTCCAGAAATACAGATGATTGGACAA 131
DB 100935 TACAGCTAATCAGAATGGCAGACACTCAACCTTCCAAAAAACAACAGATGACTGGACAA 100876
QY 132 AAGAGGATGAATCAGTGGTTGAAAGTCAATAGATTGACCAAAAACACAGGGAATTT 191
DB 100875 AAGAGGATGAATCAGTGGTCTAGAAAGTCAATAGATTGACCAAAAACACAGGGAATTT 100816
QY 192 TGACTGAAACAAGACGTGAATGGAGCAGTCTTTGAAAGTGGTTAAAAAAGAAACATCTTGTG 251
DB 100815 TGATGGCAACAAGATGTGAGTGGAGTAGTCTTTGAAAGTGGTTAAAAAAGAAATCTTGTG 100756
QY 252 ATATGGGCATCACATGGACCGACTATTCAATAGAGAGACTATTCAAGAGATTTGGGA 311
DB 100755 AAATGGGCATCAACCATGGGCAGCTATCCAAATAGAGAGCTATTCAAGAGATTTGGGA 100696
QY 312 AAACAGCCATTGAAGTTCGATTTCAGACATCTAAGATGGGAAAGCCAGCAATAAATGCTC 371
DB 100695 AAACACCCAAAGAAATCTTACTAAGACATCTAAGAGACAAAAGGCAGAAAAATATCC 100636
QY 372 CTAAGACCAAACTGTGCTCAAGAGAACCTAGAGAACTTCAAGCAAAACAAAAGG 431
DB 100635 CTAACAAACCCAGAC-----AGAAAGTAGAGAGACTTCAATCAAAACCCAGAG 100588
QY 432 GTAAAGAGAACCCAGATATGGTATATCCGTCTGCAATG----- 469
DB 100587 CTGAAGAGAACTCATGCAAGCTGACACTCTGCAAGTCCAGAGAACTCACAAA 100528
QY 470 -----AGTACAACCTGCTAAAGGTTCTAAGTCTAAGTCTAAGTCTAAG 518
DB 100527 GGACTGCAGCAAGTCCAGAGGCTAAAGGTTCTAATATCAAGAAAAAANTGACCTAGTAGA 100468
QY 519 ATAAATAGATTATCAAGAAAGGCAACCATCATAGACTGACATGTGTATCATATC 578
DB 100467 ATAAAGCAGTCCA---ACAGGATAGAGGATCTCCAGGACCAACGTCGATTCATATC 100411
QY 579 CATTTGATGAATTCAGTAAATCCATATCGTTCAAGTGGATTTAGTCTACAGCTGAAA 638
DB 100410 CTTTCAATGAGTTCAGTAAATCCATATCGTTCAAAATACATTTTCATCTACAGCTGAAA 100351
QY 639 CAGGACCCAGGCAATCTCATGATCCGATACATGATTCAAAGCCTTCAAAATACAGCAA 698
DB 100350 CAGGGCCATCAATCTCATTTGACCCAGTACATGATTTCAAGCCCTTCAAAACACAGAGA 100291
QY 699 CAGCCACAGAGAGGATGTCAAGATGAAATTTAGCAATGAGGTTTCCGATTTGCTTCAG 758
DB 100290 CAGCCACAGAGAGGACATAAAGATGAAATTTAGCAAGAGTCTTCCGATTTGCTTCAG 100231
QY 759 CTTGTATGAATTCAGTACCAATGCACTATTTCATTTTGGATGCAAGCAAAACCCATG 818
DB 100230 CTTGTATGAATTCAGTACTAATGCAACCATTTTCATTTTGGATGCAAGCAAAACCCATG 100171
QY 819 GGAAATTTGTTGGCATCAAGAGTCAACCAATGATACCAAGGAAGCCCTCTAATCAACATTTC 878
```

```
DB 100170 GAAGGATTTGGTCAGCTCACCACCGTCACTAAAGATGCCCTCATTCAGCCACTTCA 100111
QY 879 ATCTCATGATAAAACAAGTATTTTGAAGACCAATCAAGTCCAAACAAGAAAGTGCATTC 938
DB 100110 ATCTCATGATAGTCCAGTATTTTGAAGTCCATCAAGTCCAAACAAGAAAGTGCATTC 100051
QY 939 GAGAGCCAGATTTGTGGAAGTTTACTTGCCAAATAGTACTCTATCTGACAGATTTGTTA 998
DB 100050 GAGAGCCAGATTTGTGGAAGTTTACTTGCCAAATAGTACTCTATTAATTAATTTGTTA 99991
QY 999 TTAGAGTGGACATTTATCCACAGTTCTCTGAATGCCAATATGATTACTTCCAGATTTAAAA 1058
DB 99990 TTAGAGTGGATATCATTTCCACATATTCAGATTCGAAAGAGATTTCTTCCAAATAAAA 99931
QY 1059 TGCAAAATTTACAAACAACAAATATCGGAACAAAAGTAAATAATTCATATTTGTGCGAG 1118
DB 99930 TGCATTTGGCATAAGGATGAACATGCAACAAATTTCCAAATATCTCGGTCTTTGTGAGAG 99871
QY 1119 ATGGGACCGCTCTAAGGACATTTACGAAATATAAGTTGATTTTCAGAGCATTTTAAAGCAG 1178
DB 99870 ATGGTCCGAAAAAGTAAAGACATCATAGGAAATAAAGCTGATTTCAAGCTTTTAAAGTTAG 99811
QY 1179 ATTTTAAACACTGGCAGAGTCCAGAAAAGCAGAGAGAAAATTCAGAGCAAAAACAA 1238
DB 99810 ACCTTAAAGCACTGGCAGACTCTAGGAAAAGCAGAGAAAATTTCCAAAGTAAATCAG 99751
QY 1239 ATAAAAGAAAGAGAGGACCAAAAGTTGGTTTAAATATTATGACAGAAATCAAGATTTGT 1298
DB 99750 ATGGAAGCAACAGTCAAGGGGCAAAAGCTAAATGACCTGTTGACAGAAATAAGGATCTGC 99691
QY 1299 TAGATAATTTCACTATGAAACAGTACATTTCTTTGTAACAAATAAATGCCACCCAGATCAA 1358
DB 99690 TAGATAATTTCACTATGAAATGTTGATCTTTGTAACAAATAAATGTCATCCAGATCAA 99631
QY 1359 CAAAACACTTAGATTTCTTGAAGGAAATTAATGTTGCTGTATGAGATTTGATCCCTG 1418
DB 99630 CCAAACACCTAGATTTCTTGAAGGAAATTAATGTTGCTGTATGAGATTTGATCCCTG 99571
QY 1419 AGTCTAACATCAATGAGTGGTCAAGCTTACAAAGAAAGCCGAGTAGCAAACTTCACT 1478
DB 99570 AGTCTGAGAGCAATGGTGTGGTCAAGGCTTTCAAGAGGAGCCGAGTAGCAAACTTCACT 99511
QY 1479 TTCCAAAGTGTATGTAGAAACAGAAAACACACCAAAATAGAGAGATTTCTACTCTAAATC 1538
DB 99510 TTCCAAAGCTATTTAGAAAGAAAACCCACAGTAAGTAAAGAAAGATTTCCAGTCTGAATC 99451
QY 1539 TTTTACCATCAA 1549
DB 99450 TTTTATCGCAA 99440
```

RESULT 14

AC108900/c

LOCUS

DEFINITION

Pelis catus clone RP86-299K23, WORKING DRAFT SEQUENCE, 2 ordered

pieces.

AC108900

AC108900.2 GI:21954014

HTG; HTGS_PHASE2; HTGS_DRAFT.

KEYWORDS

Pelis catus (cat)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Felidae; Felis.

1 (bases 1 to 119266)

Akther, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R., Bouffard, G.G., Breen, K., Brinkley, C.,

Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,

Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,

Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,

Margulies, E.H., Masello, C., Maskeri, B., Pearson, R.,

McCluskey, J.C., McDowell, J., Paguirigan, C., Pearson, R.,

Portnoy, M.E., Prasad, A., Schuler, M.G., Stantipop, S., Thomas, J.W.,

Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, N.A.,

Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 2 (bases 1 to 119266)
 Green, E.D.
 Direct Submission
 Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 3 (bases 1 to 119266)
 Green, E.D.
 Direct Submission
 Submitted (24-JUL-2002) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 On Jul 24, 2002 this sequence version replaced gi:18464076.
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@hgri.nih.gov
 ----- Project Information
 Center project name: cfu
 Center clone name: 299K23

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 118977 bases at least Q40
 Consensus quality: 119103 bases at least Q30
 Consensus quality: 119162 bases at least Q20
 Insert size: 118000; agarose-1p
 Insert size: 119166; sum-of-contigs
 Quality coverage: 11.69x in Q20 bases; agarose-1p
 Quality coverage: 11.58x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 78597: contig of 78597 bp in length
 * 78598 78697: gap of unknown length
 * 78598 119266: contig of 40569 bp in length.

FEATURES

source

1. 119266
 /organism="Felis catus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9685"
 /clone="RP86-299K23"
 /clone_lib="RP86"

misc_feature

1. 78597
 /note="assembly_fragment
 clone end:SP6
 vector_side:left"
 78698-119266
 /note="assembly_fragment
 clone_end:17"

ORIGIN	misc_feature	vector_side:right
	Query Match	87940...119266
	Best Local Similarity	/notes="clone overlaps with GenBank Accession Number
	Matches 1148; Conservative	AC105439 clone RP86-515N8 (center project name awl)"
QY	77	GGTATCAGAATGGCAAGCAACTTAACCTTCCAGAAATACAGATGATTCGACAAAGAG 136
Db	8288	GCTATCAGAATGGCAATGCCACTGACCTACCAAGAAATACAGATGATTCGACAA 8232
QY	137	GATCTAAATCAGTGGTTAGAAAGTCATPAGATGACCAAAAACACAGGGAAATTTGACT 196
Db	8231	GATATAAATTCATGTTTGGAAAGTCATAGAATTAATCAAAAGCACAGGAAATCTTGACT 8172
QY	197	GAACAAGACGTGATGGAGCAGCTTT-GAAGTGGTTAAAAAAGAGACATCTTGTGATAT 255
Db	8171	GCAAGATATGCAATAGATTAATCTTGGAAAGTGGTTAAACGAAATCACCTTGTGATAT 8112
QY	256	GGGCATCACATGACCCAGCTATTCAATAGAGAACTATTCAAGAATTCGGGAAAC 315
Db	8111	GGGCATGACATGGATGAGCTATCCATATAGAACAGCTATTCAATAATTCAGAAAC 8052
QY	316	AGCAATTCAGATTCGATTCAGATCTAAGATGGGAAGCCAGTAATAATTCCTCTAA 375
Db	8051	ATCCTCTGAAGATCTTTAGTACATGAAAGA-GGGHAAAGGTAGTAAATAATTTCTTAA 7993
QY	376	AGACCAAACTGTGTCTCAAAAGGACGTAGAGAACTTCAAGCAAAAACAAAGGGTAA 435
Db	7992	AATACAACTTTGATGTAAAGAGAAATGGAGAACTTCAAGCAAAAACAAAGT-AT 7935
QY	436	AGGAACCCAGATATGCTTAACTCGTCTGCAATGAGTACAACTGCTAAAGTTCCTAAGTC 495
Db	7934	AGGAATTCAGTTATGGCTAATGCTCTACAGTGGTAC-----AGTC 7892
QY	496	ACTAAAAGTTGAGCTCATAGAAGATAAATAGATTTATCAAGAGAAAGGCAACCATCAT 555
Db	7891	ACTAAAAGTTGAGTTCTGGAAGATGAATAGATGACACAAAGGAAAGGCGCATCCAG 7832
QY	556	AGACTGACATGTGTATCATATCCATTTGATGAATTCAGTAATCCATATCGTTACAGTT 615
Db	7831	AGAACTGACATGTGTATCATATCTTTTGTGATGAAT-TGTAACCCATATCATACAAATC 7773
QY	616	GGATTTAGTCTACAGCTGAACAGG---ACCAGGCAATCTCATTTGATCCGATACATGA 672
Db	7772	AAATTTAGTCTACAGCTGAATAGACCCACTCAATCTCACTCATCAACAATTA 7713
QY	673	ATTCAAAGCC-----TTCACAAATACAGCAACAGCCACAGAGAGGTGTCAAGTGAAT 728
Db	7712	ATTCAAACCTTATTTGTGAAATAGAGAAACAGCTGCAGAGA-GATGTTAAGGTGAAT 7654
QY	729	TTAGCAATGAGTTTTCGATTTCTTTCAGTTCTGATG---AATTCAGTACCAATGCGAC 786
Db	7653	TTAGCAATGAGTTTTCATTTCTTTCGAGTTGTATGAAATTCACCTACTAATGCGAC 7594
QY	787	TATTCATTTGGATGCAAGCAAAACCCATGGGAAATTTGTTGGATCAAAAGTCACAA 846
Db	7593	CATTCAATTTGGAGTTAAAGGCAAAACCCATGGAATAATTTCTTTACATGAAGTCAACAA 7534
QY	847	TGATACCAAGCAAGCCCTCATTAACCAATTTCAATCTGATGATAAACAAGTATTTGAAGA 906
Db	7533	TTTACCAGGAGAGCCCTCATTTGACCAATTTAATGATGATCCACAGTATTTGAAGA 7474
QY	907	CCATCAAGTCCAAACAAAGAGAGTGCATTCGAGAGCCAGATTTGTGGAAGTTTACT 966
Db	7473	CCATCAGGTCCAAAAGCAAGAGAGTACATTTGAGAGCCAGATTTGTAGACGTTTACT 7414
QY	967	GCCAAATAGTACTCTCTCTGACAGATTTGTTTATTGAAGTGCACATTTATCCACAGTTCTC 1026
Db	7413	GCTAAACCGTAGGTATCTCTGACAGATTTG-TATTGAAGCAAGGTCAATACAAAGTACTC 7355


```
/note="clone overlaps with GenBank Accession Number
AC110666 clone RP81-156E11 (center project name avv)"
misc_feature
196724..225885
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature
223792..225885
/note="clone overlaps with GenBank Accession Number
AC110671 clone RP81-69D24 (center project name avv)"
ORIGIN
Query Match 49.0%; Score 759; DB 2; Length 225885;
Best Local Similarity 77.2%; Pred. No. 8.8e-134;
Matches 1143; Conservative 0; Mismatches 285; Indels 53; Gaps 16;
77 GGTATCAGAAATGGCAAGCACTTAACCTTCAGAAAATACAGATGATTGGCAAAAGAG 136
Db GCTATCAGAAATGACAGCACATGAACTTGCCTGCAAAATTAACATGATGGCAAAAGAA 125839
QY 137 GATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAACACAGGAAATTTTGACT 196
Db GATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAATTAACATGATGGCAAAAGAA 125839
QY 197 GAACAGAGCGTGAATGGACGCTCTGAGTGGTTAAAGAAAGACATCTGTTGATATG 256
Db GATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAATTAACATGATGGCAAAAGAA 125839
QY 125778 GCAGAAATGACATGGAGCAATCTTGAAGTGGT-----AAAATCACCCTTTGTTGATTG 125724
Db GATGTAATCAGTGGTTAGAAAGTCATAAGATTGACCAAAATTAACATGATGGCAAAAGAA 125839
QY 257 GGCATCACATCGGACCACTTCAAAATAGAAAGTATTTCAAAAGAAATTCGGGAAACA 316
Db GGCATCACATCGGACCACTTCAAAATAGAAAGTATTTCAAAAGAAATTCGGGAAACA 316
QY 317 GCATATGAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAA 376
Db GCATATGAGATTCGATTCAGACATCTAAGATGGGAAAGCCAGTAAATGCTCTTAAA 376
QY 125677 TCTTCTGAAGATCCCTATTGAGACATGTAAGATGAAAGAG-GTAGTAAATGTTCTCTTAA 125619
Db TCTTCTGAAGATCCCTATTGAGACATGTAAGATGAAAGAG-GTAGTAAATGTTCTCTTAA 125619
QY 377 GACCAACTGTCTCTCAAAAGGAAGCTAGAGAACTTCAAAAGCAAAACAAAGCGTAAA 436
Db GACCAACTGTCTCTCAAAAGGAAGCTAGAGAACTTCAAAAGCAAAACAAAGCGTAAA 436
QY 437 GAGAACCCAGATATGGCTAATCCGCTGCAATGAGTACACACTGCTAAAGGTTCTAAGTCA 496
Db GAGAACCCAGATATGGCTAATCCGCTGCAATGAGTACACACTGCTAAAGGTTCTAAGTCA 496
QY 125558 GAGAATACGCTATAGCTAATGCCCTCAATGAGTACAGTACTTAAAGGTTCTCAGCCA 125499
Db GAGAATACGCTATAGCTAATGCCCTCAATGAGTACAGTACTTAAAGGTTCTCAGCCA 125499
QY 497 CTAAAGTTGAGCTCATAGAGATATAAGATTTATCAAAAGGAAGGCAACCACTCCATA 556
Db CTAAAGTTGAGCTCATAGAGATATAAGATTTATCAAAAGGAAGGCAACCACTCCATA 556
QY 125498 CTAAAGTTGAGCTCATAGAGATATAAGATTTATCAAAAGGAAGGCAACCACTCCATA 125439
Db CTAAAGTTGAGCTCATAGAGATATAAGATTTATCAAAAGGAAGGCAACCACTCCATA 125439
QY 557 GACCTGACATGTGATATCATATTCATTTGATGAAATTCAGTAACTCATATCGTTACAGTTG 616
Db GACCTGACATGTGATATCATATTCATTTGATGAAATTCAGTAACTCATATCGTTACAGTTG 616
QY 125438 GAATTAACATGTGATATCATATTCATTTGATGAAATTCAGTAACTCATATCGTTACAGTTG 125381
Db GAATTAACATGTGATATCATATTCATTTGATGAAATTCAGTAACTCATATCGTTACAGTTG 125381
QY 617 GATTTTAGTCTACAGCCTGAAACAGG-ACCAGGCAATCTCATTTGATCCGATCATGAA 673
Db GATTTTAGTCTACAGCCTGAAACAGG-ACCAGGCAATCTCATTTGATCCGATCATGAA 673
QY 125380 AGTTTGTGCTGATCTGAAACAGGCAACCACTCACTTATTTGATCCATCATGAA 125321
Db AGTTTGTGCTGATCTGAAACAGGCAACCACTCACTTATTTGATCCATCATGAA 125321
QY 674 TTCAGAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGC 733
Db TTCAGAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGC 733
QY 125320 TTCAGAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGC 125270
Db TTCAGAGCCTTCACAAATACAGCAACAGCCACAGAGAGGATGTCAAGATGAAATTTAGC 125270
QY 734 AATGAGGTTTTCGATTTGCTTTCAGCTTGTATG- AATTCAGTACCAATGGCAGCTATTCA 792
Db AATGAGGTTTTCGATTTGCTTTCAGCTTGTATG- AATTCAGTACCAATGGCAGCTATTCA 792
QY 125269 AA-CAAAATTTTCAATTTGCTTTCAGCTTGTATGAAATTTCACTACTAATGGCAGCTTCA 125211
Db AA-CAAAATTTTCAATTTGCTTTCAGCTTGTATGAAATTTCACTACTAATGGCAGCTTCA 125211
QY 793 TTTTGGAGTCAAGAGCAAAACCCATGGGAAATTTTGGCATCAAAAGTCAAAATGATPAC 852
Db TTTTGGAGTCAAGAGCAAAACCCATGGGAAATTTTGGCATCAAAAGTCAAAATGATPAC 852
QY 853 CAAGGAGCCCTCATTAACCATTTCAATCTGATGATTAACAGATATTTTGAAGCCATCA 912
Db CAAGGAGCCCTCATTAACCATTTCAATCTGATGATTAACAGATATTTTGAAGCCATCA 912
QY 125150 CAAGGAGCCCTCATTAACCATTTCAATCTGATGATTAACAGATATTTTGAAGCCATCA 125091
Db CAAGGAGCCCTCATTAACCATTTCAATCTGATGATTAACAGATATTTTGAAGCCATCA 125091
QY 913 AGTCCAAAGCAAGAGAGAGTGCATTCGAGAGCAAGATTTTGGAGGTTTCTACGCCAAA 972
Db AGTCCAAAGCAAGAGAGAGTGCATTCGAGAGCAAGATTTTGGAGGTTTCTACGCCAAA 972
```

Search completed: April 25, 2004, 18:58:31

Job time : 4090.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 25, 2004, 07:00:48 ; Search time 763.869 Seconds
(without alignments)
16458.288 Million cell updates/sec

Title: US-10-051-835-16
Perfect score: 421
Sequence: 1 atggagttcttcattctg.....ttctatgtctcttcagcgag 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_hcc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_hcc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: em_gss_hum:**
- 18: em_gss_inv:**
- 19: em_gss_pln:**
- 20: em_gss_vrt:**
- 21: em_gss_fun:**
- 22: em_gss_mam:**
- 23: em_gss_mus:**
- 24: em_gss_pro:**
- 25: em_gss_rod:**
- 26: em_gss_phg:**
- 27: em_gss_vrl:**
- 28: gb_gss1:**
- 29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	169	40.1	961	10 BE796744	BE796744 601587368
2	68.8	16.3	533	28 AQ418633	AQ418633 RPCI-11-2
3	58.4	13.9	686	10 BE901475	BE901475 601674934
4	56.6	13.4	447	12 BT013264	BT013264 PM4-ET015

C	5	56.6	13.4	598	9	AL549589
C	6	56.6	13.4	635	13	BX504960
C	7	56.6	13.4	766	12	BT460285
C	8	56.6	13.4	906	14	CF994492
C	9	56.6	13.4	990	13	EX334700
C	10	55.6	13.2	498	28	AQ624094
C	11	55.6	13.2	518	28	AQ145903
C	12	55.6	13.2	765	28	BZ600533
C	13	55.6	13.2	872	12	BI911783
C	14	55.6	13.1	640	28	AQ108606
C	15	55.6	13.1	645	28	AQ077228
C	16	55.6	13.1	676	9	AV730039
C	17	55.6	13.1	682	9	AV731500
C	18	54.6	13.0	495	28	AQ235894
C	19	54.6	13.0	554	9	AL705106
C	20	54.6	13.0	795	12	BG698619
C	21	54.6	13.0	930	13	BQ689079
C	22	54.2	12.9	365	28	B30587
C	23	54.2	12.8	214	10	BE142974
C	24	54.2	12.8	339	28	AQ566892
C	25	54.2	12.8	341	28	AQ080641
C	26	54.2	12.8	348	10	BF920788
C	27	54.2	12.8	408	28	B33662
C	28	53.4	12.7	401	10	BF759523
C	29	53.4	12.7	522	12	BG703419
C	30	53.4	12.7	532	12	EM509480
C	31	53.2	12.6	620	28	AQ586844
C	32	53.2	12.6	473	10	AW876755
C	33	53.2	12.6	517	28	B49322
C	34	53.2	12.6	590	10	BE144970
C	35	53.2	12.6	727	14	CF994332
C	36	52.8	12.5	651	12	BG718914
C	37	52.8	12.5	701	12	B559378
C	38	52.8	12.5	833	13	BQ225374
C	39	52.6	12.5	320	10	BE143157
C	40	52.6	12.5	360	12	BG981236
C	41	52.6	12.5	478	13	BE112097
C	42	52.6	12.5	483	28	B69221
C	43	52.6	12.5	762	28	B90054
C	44	52.4	12.4	275	28	B58637
C	45	52.4	12.4	403	28	AQ278768

ALIGNMENTS

RESULT 1
BE796744
LOCUS
DEFINITION
601587368F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3941440 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 961)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCW793 row: h column: 17
High quality sequence start: 25
High quality sequence stop: 689.

BE796744 961 bp mRNA linear EST 20-SEP-2000
601587368F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3941440 5',
mRNA sequence.

ACCESSION BE796744.1 GI:10217942

VERSION BE796744

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 961)

NIH-MGC http://mgc.nci.nih.gov/

AUTHORS

TITLE

JOURNAL

COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCW793 row: h column: 17

High quality sequence start: 25

High quality sequence stop: 689.

```

FEATURES
  source
    Location/Qualifiers
      1..961
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:3941440"
        /tissue_type="small cell carcinoma"
        /cell_line="MGC3"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_7"
        /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
        EcoRI; cDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGACGAG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using ZAP-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)."
```

ORIGIN

```

Query Match      40.1%; Score 169; DB 10; Length 961;
Best Local Similarity 92.2%; Pred. No. 5.4e-37;
Matches 178; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 18 TGGTAGTTCGGTCTCTCTGCTCAGATGAGCTGTAGACTCTGGCATCTGT 77
DB 110 TGGTGGGTTCCTGCTCTCACTGACTTCAAGAATGAAGCCCGACCTCGCAATGCTGT 169

QY 78 ATGAATCTTCCTTAGAATCAACCCATTTGTGAACACCAACATTAATAAGAGCAGTATCTT 137
DB 170 ACGAATCTTCCTTAGAATCAACCCATTTGTGAACACCAACATTAATAAGAGCAGTATCTT 229

QY 138 TGAAGATTGGAGAGTCACCAAGAGTCCCACTGGCTCCCTACAAAGTTTATTATGTGAG 197
DB 230 TGAAGATTGGAGAGTCACCAAGAGTCCCACTGGCTCCCTACAAAGTTTATTATGTGAG 289

QY 198 GACTCGTCTTCAG 210
DB 290 GACTCGTCTTCAG 302

RESULT 2
AQ418633
LOCUS
DEFINITION
  AQ418633 533 bp DNA linear GSS 23-WAR-1999
  genomic survey sequence.
ACCESSION
  AQ418633
VERSION
  AQ418633.1 GI:4476357
SOURCE
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 533)
  Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
  Venter,J.C.
  Map Building
  Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
  Unpublished (1997)
  Other GSSs: RPCI-11-20304.TJ
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genet cs (info@resgen.com). BAC end search page:
  http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
```

```

Seq primer: T7
Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1..533
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="GDB:7577907"
        /db_xref="taxon:9606"
        /clone="RPCI-11-20304"
        /sex="Male"
        /cell_type="Lymphocytes"
        /clone_lib="RPCI-11"
        /note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
        RPCI11 Human Male BAC Library"
```

ORIGIN

```

Query Match      16.3%; Score 68.8; DB 28; Length 533;
Best Local Similarity 97.2%; Pred. No. 1.3e-08;
Matches 70; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGATTTCTTCATTTCTGTTAGTTCGTGGTCTCTCTCGCTTCAGGAATGAAGCTGTAG 60
DB 457 ATGAGATTTCTTCATTTCTGTTAGTTCGTGGTCTCTCTCGCTTCAGGAATGAAGCTGTAG 516

QY 61 AACTCTCGCATG 72
DB 517 AACTCTCGCATG 528

RESULT 3
BE901475/c
LOCUS
DEFINITION
  BE901475 686 bp mRNA linear EST 29-SEP-2000
  mRNA sequence.
ACCESSION
  BE901475
VERSION
  BE901475.1 GI:10390693
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 686)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-rc@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Ling Hong/Rubin Laboratory
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
  Plate: LLCM835 row: d column: 09
  High quality sequence start: 3
  High quality sequence stop: 686.
  Location/Qualifiers
    1..686
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:3957464"
      /tissue_type="choriocarcinoma"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_21"
      /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
      Site 2: EcoRI; cDNA made by oligo-dT priming.
      Directionally cloned into EcoRI/XhoI sites using the
      following 5' adaptor: GGACGAG(G). Size-selected >500bp
      for average insert size 1.8kb. Library constructed by
      Ling Hong in the laboratory of Gerald M. Rubin (University
      of California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies)."
```

ORIGIN

Query Match 13.9%; Score 58.4; DB 10; Length 686;
 Best Local Similarity 85.5%; Pred. No. 1.2e-05;
 Matches 65; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 3 GGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
 DB 258 GGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 199
 QY 63 CTCTCGGATGCTGTGTTA 78
 DB 198 CTTTGGGTGATGTTA 183

RESULT 4

BI013264/c
 LOCUS PM4-ET0154-200101-009-c09 ET0154 Homo sapiens linear EST 13-JUN-2001
 DEFINITION PM4-ET0154-200101-009-c09 ET0154 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BI013264
 VERSION BI013264.1 GI:14417335
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 447)
 AUTHORS Dias Neto E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 2002663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&t2=PM4-ET0154-
 200101-009-c09&t3=2001-01-20&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 446.

FEATURES

source
 1..447
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="ET0154"
 /note="Organ: lung tumor; Vector: puc18; Site: 1: SmaI;
 Site 2: SmaI; A mini-library was made by cloning products
 derived from ORSTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 13.4%; Score 56.6; DB 12; Length 447;
 Best Local Similarity 87.3%; Pred. No. 3.5e-05;
 Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 61

ORIGIN

Db 228 TGGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 169
 QY 62 ACTCTCGGATG 72
 Db 168 CTTTGGCAGTG 158

RESULT 5

AL549589/c
 LOCUS AL549589 Homo sapiens linear EST 31-MAY-2003
 DEFINITION AL549589 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 clone CS0DI048YC03 5-PRIME, mRNA sequence.
 ACCESSION AL549589
 VERSION AL549589.2 GI:31271407
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 598)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12885723.
 Contact: Genoscope
 Genoscope Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Paraday Avenue Genoscope sequence ID : CS0DI048AB02QPl.

FEATURES

source
 1..598
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI048YC03"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dt)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 13.4%; Score 56.6; DB 9; Length 598;
 Best Local Similarity 87.3%; Pred. No. 3.7e-05;
 Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 61

Db 229 TGGAGTTTCTTCATTCTGGTAGTTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 170

QY 62 ACTCTCGGATG 72

Db 169 TCTTCCGGTG 159

RESULT 6

BX504960
 LOCUS BX504960 Homo sapiens linear EST 04-SEP-2003
 DEFINITION BX504960 Homo sapiens hicc3 Homo sapiens cDNA clone
 DKF2p686I13187 5', mRNA sequence.
 ACCESSION BX504960
 VERSION BX504960.1 GI:32032507
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 635)

AUTHORS
Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R.,
Meves, H.W., Well, S., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.

TITLE
EST (Poustka, A., Albert, R., Moosmayer, P., Schupp, I.,
Wellenreuther, R., et al.)

JOURNAL
Unpublished (2003)

COMMENT
Contact: MIPS
Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686l13187) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Reubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
Location/Qualifiers
1..635
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686l13187"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/notes="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN
Query Match 13.4%; Score 56.6; DB 13; Length 635;
Best Local Similarity 87.3%; Pred. No. 3.8e-05;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGAGTTTCTTCATTCTGGTAGGTTCTGCTCTCTGCTTCAGGAATGAGCTGTAGA 61
Db |||||
165 TGGAGTTTATCTCTCTGGTAGGTTCTGCTCTCTGCTTCAGGAGTGAAGCTGCAGA 224
QY 62 ACTCTGGGATG 72
Db |||||
225 TCTTGGCGGTG 235

RESULT 7
BI460285/c
LOCUS 603201916F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267955 5',
DEFINITION mRNA sequence.
ACCESSION BI460285
VERSION BI460285.1 GI:15250941
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 766)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11675 row: p column: 04
High quality sequence stop: 701.

FEATURES
Location/Qualifiers
1..766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5267955"
/lab_host="DH10B"
/clone_lib="NIH_MGC_97"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
(gtccag); Oligo-dT primed using primer
5'-TTTTTTTTTTTTTTN-3', size-selected for average
insert size 2.2 kb and normalized to 10^5. This is a
primary library enriched for full-length clones and
constructed using the Cap-trapper method (Carninci, in
preparation). Library constructed by M. Brownstein
(NIH/NHGRI, National Institutes of Health). Note: this is
a NIH_MGC Library."

ORIGIN
Query Match 13.4%; Score 56.6; DB 12; Length 766;
Best Local Similarity 87.3%; Pred. No. 4e-05;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGAGTTTCTTCATTCTGGTAGGTTCTGCTCTCTGCTTCAGGAATGAGCTGTAGA 61
Db |||||
242 TGGAGTTTCTTCATTCTGGTAGGTTCTGCTCTCTGCTTCAGGAGTGAAGCTGCAGA 183
QY 62 ACTCTGGGATG 72
Db |||||
182 CTTTCGCGGTG 172

RESULT 8
CF994492/c
LOCUS CF994492 906 bp mRNA linear EST 25-NOV-2003
DEFINITION AGENCOURT 15621259 NIH_MGC_147 Homo sapiens cDNA clone
IMAGE:30520429 5', mRNA sequence.
ACCESSION CF994492
VERSION CF994492.1 GI:38510552
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaps@remail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM593 row: i column: 14
High quality sequence stop: 564.

FEATURES
Location/Qualifiers
1..906
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30520429"
/tissue_type="Human Placenta"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC_147"

/note="Organ: placenta; Vector: pBluescriptB; Site_1: all.XhoI; Site_2: BamH; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: This is a NIH_MGC library."

ORIGIN

Query Match 13.4%; Score 56.6; DB 14; Length 906;
Best Local Similarity 87.3%; Pred. No. 4.1e-05;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGAGTTTCTTCATCTCGTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGTAGA 61
|||||
Db 211 TGGAGTTTATTCCTCTCTGTTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGCAGA 152
|||||

QY 62 ACTCTGCGATG 72
|||||

Db 151 TCTTTGCGGTG 141
|||||

RESULT 9

EX334700/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

EX334700 990 bp mRNA linear EST 01-MAY-2003
clone CSODI005YJ22 5-PRIME, mRNA sequence.

EX334700
EX334700.1 GI:30308308
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 990)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9886.f,
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CSODI005D811QPI.

FEATURES

Location/Qualifiers
1..990
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI005YJ22"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 13.4%; Score 56.6; DB 13; Length 990;
Best Local Similarity 87.3%; Pred. No. 4.2e-05;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGGAGTTTCTTCATCTCGTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGTAGA 61
|||||

Db 234 TGGAGTTTATTCCTCTCTGTTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGCAGA 175
|||||

QY 62 ACTCTGCGATG 72
|||||

Db 174 TCTTTGCGGTG 164
|||||

RESULT 10
AC624094/c
LOCUS

AC624094 498 bp DNA linear GSS 16-JUN-1999
HS 2100 A1 B10 MR CIT Approved Human Genomic Sperm Library D Homo
sapiens Genomic clone Plate=2100 Col=19 Row=C, genomic survey
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC624094
AC624094.1 GI:5086486
GSS.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 498)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
1049764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2100 row: C column: 19
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 498.

FEATURES
source

Location/Qualifiers
1..498
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2100 Col=19 Row=C"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelosAC11; BAC Clones in
E-Coli DH10B"

ORIGIN

Query Match 13.2%; Score 55.6; DB 28; Length 498;
Best Local Similarity 87.1%; Pred. No. 6.9e-05;
Matches 61; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 GGAGTTTCTTCATCTCGTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGTAGA 62
|||||

Db 272 GGAGTTTCTTCCTCTCTGTTAGTTCGGTCTCTCTCGCTTCAGGAATGAAGCTGCAGAC 213
|||||

QY 63 CTCTGCGATG 72
|||||

Db 212 CTCCTGGTGG 203
|||||

RESULT 11
AQ145903/c

AQ145903 518 bp DNA linear GSS 26-SEP-1998
HS 2219_B2_G02_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens Genomic clone Plate=2219 Col=4 Row=N, genomic survey
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AQ145903
AQ145903.1 GI:3536556
GSS.
Homo sapiens (human)

and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 13.1%; Score 55.6; DB 12; Length 872;
Best Local Similarity 87.1%; Pred. No. 7.9e-05; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 9;

QY 3 GGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 62
DB 649 GGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 590

QY 63 CTCCTCGCATG 72
DB 589 CTTCCGGGTG 580

RESULT 14

LOCUS AQ108606/c 640 bp DNA linear GSS 29-AUG-1998
DEFINITION CIT-HSP-2379119.TF CIT-HSP Homo sapiens genomic clone 2379119,
genomic survey sequence.
ACCESSION AQ108606
VERSION AQ108606.1 GI:3485296
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 640)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building

JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2379119.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES

Location/Qualifiers
1..640
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2379119"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 13.1%; Score 55; DB 28; Length 640;
Best Local Similarity 85.9%; Pred. No. 0.00011; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 10;

QY 2 TGGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 61
DB 206 TGGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 147

QY 62 ACTCTGCGATG 72
DB 146 CCTTCTGGGTG 136

RESULT 15

LOCUS AQ077228/c 645 bp DNA linear GSS 20-AUG-1998
DEFINITION CIT-HSP-2359F6.TR CIT-HSP Homo sapiens genomic clone 2359F6,
genomic survey sequence.

ACCESSION AQ077228
VERSION AQ077228.1 GI:3438412
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 645)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building

JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

Location/Qualifiers
1..645
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2359F6"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 13.1%; Score 55; DB 28; Length 645;
Best Local Similarity 85.9%; Pred. No. 0.00011; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 10;

QY 2 TGGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 61
DB 275 TGGAGTTTCTTCATTCTGCTAGGTTCTGCTCTCTGCTTTCAGGAATGAAGCTGTAGA 216

QY 62 ACTCTGCGATG 72
DB 215 CCTTCTGGGTG 205

Search completed: April 26, 2004, 01:59:26
Job time : 767.869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 20.8853 Seconds
(without alignments)
11186.557 Million cell updates/sec

Title: US-10-051-835-16
Perfect score: 421
Sequence: 1 atggagtttcttcattctgg.....ttctatgtccttcagag 421

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq*
5: /cgn2_6/ptodata/2/ina/PCRUS COMB.seq*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
C 1	55	13.1	426	4	US-09-621-976-10375
C 2	54.8	13.0	501	4	US-09-621-976-10548
C 3	52.8	12.5	579	4	US-09-621-976-1382
C 4	51.6	12.3	356	4	US-09-621-976-9793
C 5	51.6	12.3	391	4	US-09-621-976-7956
C 6	51.6	12.3	466	4	US-09-621-976-8468
C 7	51.6	12.3	481	4	US-09-621-976-8002
C 8	51.6	12.3	481	4	US-09-621-976-8003
C 9	49.8	11.8	418	4	US-09-621-976-15076
C 10	49.8	11.8	468	4	US-09-621-976-15059
C 11	49.8	11.8	486	4	US-09-621-976-14952
C 12	49.8	11.8	517	4	US-09-621-976-10428
C 13	49.8	11.8	80246	3	US-09-078-294-4
C 14	49.8	11.8	80595	3	US-09-078-294-3
C 15	47	11.2	443	4	US-09-621-976-10565
C 16	47	11.2	510	4	US-09-621-976-17134
C 17	46.2	11.0	398	4	US-09-621-976-10095
C 18	46	10.9	341	4	US-09-621-976-10120
C 19	45.4	10.8	432	4	US-09-621-976-12460
C 20	45.2	10.7	451	4	US-09-621-976-10352
C 21	45.2	10.7	496	4	US-09-621-976-10365
C 22	45	10.7	421	4	US-09-621-976-10768
C 23	45	10.7	488	4	US-09-621-976-986
C 24	45	10.7	494	4	US-09-621-976-635
C 25	44.8	10.6	2381	2	US-08-736-770-4
C 26	44.6	10.6	315	4	US-09-621-976-9027
C 27	44.6	10.6	433	4	US-09-621-976-10177

C 28	44.6	10.6	440	4	US-09-621-976-12535	Sequence 12535, A
C 29	44.6	10.6	458	4	US-09-621-976-12564	Sequence 12564, A
C 30	44.6	10.6	488	4	US-09-621-976-11085	Sequence 11085, A
C 31	44.6	10.6	510	4	US-09-621-976-9389	Sequence 9389, Ap
C 32	44.6	10.6	1043	4	US-09-422-576D-6	Sequence 6, Appl1
C 33	44.6	10.6	1091	4	US-09-422-576D-5	Sequence 5, Appl1
C 34	44.2	10.5	482	4	US-09-621-976-1994	Sequence 1994, Ap
C 35	43.6	10.4	128779	4	US-09-497-855A-38	Sequence 38, Appl1
C 36	43	10.2	430	4	US-09-621-976-10289	Sequence 10289, A
C 37	43	10.2	442	4	US-09-621-976-16336	Sequence 16336, A
C 38	43	10.2	512	4	US-09-621-976-10927	Sequence 10927, A
C 39	43	10.2	513	4	US-09-621-976-16337	Sequence 16337, A
C 40	43	10.2	572	4	US-09-621-976-10590	Sequence 10590, A
C 41	42.8	10.2	418	4	US-09-621-976-18235	Sequence 18235, A
C 42	42.6	10.1	446	4	US-09-621-976-18867	Sequence 18867, A
C 43	42.6	10.1	1165	4	US-09-422-576D-25	Sequence 25, Appl1
C 44	42.6	10.1	1831	4	US-09-422-576D-1	Sequence 1, Appl1
C 45	42	10.0	375	4	US-09-621-976-10776	Sequence 10776, A

ALIGNMENTS

RESULT 1
US-09-621-976-10375/c
; Sequence 10375, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10375
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10375

Query Match 13.1%; Score 55; DB 4; Length 426;
Best Local Similarity 77.1%; Pred. No. 1e-08; 17; Indels 0; Gaps 0;
Matches 64; Conservative 2; Mismatches 17

QY	4	GAGTTTCTTCATCTGCTAGTTCGTCGTCCTCTCTCTGCTTCAGGAATCAGCTGTAGAAC	63
Db	305	GAGTTTCTTCATCTGCTAGTTCGTCGTCCTCTCTCTGCTTCAGGAATCAGCTGTAGAAC	246
QY	64	TTTGGGATGCTGTTATGACATTC	86
Db	245	TTTGGGATGCTGTTATGACATTC	223

RESULT 2
US-09-621-976-10548/c
; Sequence 10548, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10548
; LENGTH: 501
; TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-621-976-10548
Query Match      13.0%; Score 54.8; DB 4; Length 501;
Best Local Similarity 79.3%; Pred. No. 1.3e-08;
Matches 65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4 GAGTTCCTTCATCTCGTAGGTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAAC 63
Db 305 GAGTTCCTTCCTCTCGTGGTTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAAC 246

QY 64 TCTGCGATGCTGTATGAACCT 85
Db 245 TTTCGCGTGAGTGTACAGCTT 224

RESULT 3
US-09-621-976-1382/c
; Sequence 1382, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1382
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 40..246
; NAME/KEY: sig_peptide
; LOCATION: 40..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 9.5
; OTHER INFORMATION: seq FVLLFVLSLAA/AH
US-09-621-976-1382

Query Match      12.5%; Score 52.8; DB 4; Length 579;
Best Local Similarity 83.3%; Pred. No. 6.5e-08;
Matches 60; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGGAGTTCTTCATCTCGTAGGTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAG 60
Db 303 ATGGAGTTCTTCCTCTCGTAGGTCGTGGTCTCTCGCTTCAGGATGAAGCTGTAG 244

QY 61 AACTCTCGGATG 72
Db 243 ACCTTTGGGTG 232

RESULT 4
US-09-621-976-9793/c
; Sequence 9793, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 9793
; LENGTH: 356

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-9793
Query Match      12.3%; Score 51.6; DB 4; Length 356;
Best Local Similarity 76.8%; Pred. No. 1.3e-07;
Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GGAGTTCTTCATCTCGTAGGTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAA 62
Db 267 GGAGTTCTTCCTCTCGTGGTTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAA 208

QY 63 CTCTGCGATGCTGTATGAACCT 84
Db 207 CTTCGCGTGCTGTATGAACCT 186

RESULT 5
US-09-621-976-7956/c
; Sequence 7956, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7956
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 306,313,315
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-7956

Query Match      12.3%; Score 51.6; DB 4; Length 391;
Best Local Similarity 76.8%; Pred. No. 1.3e-07;
Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GGAGTTCTTCATCTCGTAGGTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAA 62
Db 284 GGAGTTCTTCCTCTCGTGGTTCGTGGTCTCTCTCGCTTCAGGATGAAGCTGTAGAA 225

QY 63 CTCTGCGATGCTGTATGAACCT 84
Db 224 CTTCGAGTGAGTGTATGAACCT 203

RESULT 6
US-09-621-976-8468/c
; Sequence 8468, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8468
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8468
```

Query Match 12.3%; Score 51.6; DB 4; Length 466;
Best Local Similarity 76.8%; Pred. No. 1.5e-07;
Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
DB 284 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 225

QY 63 CTCTGGGATGCTGTTATGAAC 84
DB 224 GTTCGAGGTGAGTGTACAACT 203

RESULT 7
US-09-621-976-8002/c
; Sequence 8002, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8002
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-8002

Query Match 12.3%; Score 51.6; DB 4; Length 481;
Best Local Similarity 76.8%; Pred. No. 1.5e-07;
Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
DB 284 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 225

QY 63 CTCTGGGATGCTGTTATGAAC 84
DB 224 GTTCGAGGTGAGTGTACAACT 203

RESULT 8
US-09-621-976-8003/c
; Sequence 8003, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8003
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 76
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-8003

Query Match 12.3%; Score 51.6; DB 4; Length 481;
Best Local Similarity 76.8%; Pred. No. 1.5e-07;

Matches 63; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 62
DB 284 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 225

QY 63 CTCTGGGATGCTGTTATGAAC 84
DB 224 GTTCGAGGTGAGTGTACAACT 203

RESULT 9
US-09-621-976-15076/c
; Sequence 15076, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15076
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15076

Query Match 11.8%; Score 49.8; DB 4; Length 418;
Best Local Similarity 82.6%; Pred. No. 5.6e-07;
Matches 57; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 63
DB 401 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 342

QY 64 TCTGCGATG 72
DB 341 CTTGTGGTG 333

RESULT 10
US-09-621-976-15059/c
; Sequence 15059, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15059
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15059

Query Match 11.8%; Score 49.8; DB 4; Length 468;
Best Local Similarity 82.6%; Pred. No. 5e-07;
Matches 57; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 63
DB 401 GGAGTTTCTTCATCTCGGTAGGTCGTGCTCTCTCTGGCTTCAGGAATGAAGCTGTAGAA 342

QY 64 TCTGCGATG 72

```
Db      341 CTTGTGGTG 333

RESULT 11
US-09-621-976-14952/c
; Sequence 14952, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14952
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 119
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-14952

Query Match      11.8%; Score 49.8; DB 4; Length 486;
Best Local Similarity 82.6%; Pred. No. 6.1e-07;
Matches 57; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      4 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63
Db      401 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63

QY      64 TCTGCGATG 72
Db      341 CTTGTGGTG 333

RESULT 12
US-09-621-976-10428/c
; Sequence 10428, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10428
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10428

Query Match      11.8%; Score 49.8; DB 4; Length 517;
Best Local Similarity 74.1%; Pred. No. 6.3e-07;
Matches 63; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY      4 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63
Db      384 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63

QY      64 TCTGCGATGCTTATGAATCTCC 88
Db      324 TTCAGCATGAGTGTACAGCTCTC 300
```

```
RESULT 13
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match      11.8%; Score 49.8; DB 3; Length 80246;
Best Local Similarity 82.6%; Pred. No. 8.5e-06;
Matches 57; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      4 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63
Db      59664 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63

QY      64 TCTGCGATG 72
Db      59604 TTCGCAGTG 59596

RESULT 14
US-09-078-294-3/c
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match      11.8%; Score 49.8; DB 3; Length 80595;
Best Local Similarity 82.6%; Pred. No. 8.5e-06;
Matches 57; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      4 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63
Db      59897 GAGTTCTTCATCTCTGAGTTCGGTCTCTCTGCTTCAGGAATGAGCTGTAGAAC 63

QY      64 TCTGCGATG 72
Db      59837 TTCGCAGTG 59829

RESULT 15
US-09-621-976-10565/c
; Sequence 10565, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
```

```

; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10565
; LENGTH: 443
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 336
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-10565

```

```

Query Match      11.2%; Score 47; DB 4; Length 443;
Best Local Similarity 78.9%; Pred. No. 5e-06;
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      2 TGGAGTTTCATCTCGTAGGTTCTGCTGCTCTCAGGATGAGCTCTAGA 61
      285 TGGAGTTTCATCTCGTAGGTTCTGCTGCTCTCAGGATGAGCTCTAGA 61
      62 ACTCTGGGATG 72
      225 CCTTCGAGTG 215
Db

```

Search completed: April 26, 2004, 02:11:09
Job time : 21.8853 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 24, 2004, 23:47:42 ; Search time 111.862 Seconds
(without alignments)
15988.422 Million cell updates/sec

Title: US-10-051-835-16
Perfect score: 421
Sequence: 1 atggagttcttcattctgg.....ttctatgccttcagcgag 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_25Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	421	100.0	421	8	ACD42206 Human HLA
2	160	38.0	481	8	ACH32547 Human end
3	141	33.5	415	4	Aai93591 Human pol
4	56.6	13.4	567	5	AAS90405 DNA encod
5	56.6	13.4	2566	4	AaH18588 Human cDN
6	56.4	13.4	339	5	AAS66944 DNA encod
7	56.4	13.4	648	5	AAS78088 DNA encod
8	56.4	13.4	648	5	AAS99070 DNA encod
9	56.4	13.4	648	5	AAS64261 DNA encod
10	55	13.1	380	3	AAC05498 Human sec
11	55	13.1	1173	5	AAS99076 DNA encod
12	54.8	13.0	465	4	Aai19816 Probe #97
13	54.8	13.0	465	4	Aai10563 Probe #49
14	54.8	13.0	465	4	ABA52210 Human foe
15	54.8	13.0	465	4	ABA64843 Human foe
16	54.8	13.0	465	4	Aai131816 Probe #50
17	54.8	13.0	465	4	Aai45011 Probe #13
18	54.8	13.0	465	4	ABA46982 Human bra
19	54.8	13.0	465	4	ABA22015 Probe #48
20	54.8	13.0	465	4	ABA31959 Probe #10
21	54.8	13.0	465	4	Aak25933 Human bon
22	54.8	13.0	465	4	Aak39006 Human bon
23	54.8	13.0	465	4	Aak13272 Human bra

C	24	54.8	13.0	465	4	AAK00486 Human bra
C	25	54.8	13.0	465	4	ABs38593 Human liv
C	26	54.8	13.0	465	4	ABs25522 Human liv
C	27	54.8	13.0	465	5	Aai00495 Probe #48
C	28	54.8	13.0	465	5	Aai05532 Probe #55
C	29	54.8	13.0	465	6	ABs13090 Human gen
C	30	54.8	13.0	465	6	ABs00516 Human gen
C	31	54.6	13.0	2943	5	AA80270 DNA encod
C	32	54.2	12.9	558	5	AA864258 DNA encod
C	33	54.2	12.9	1227	5	AA872945 DNA encod
C	34	54	12.8	552	5	AA891453 DNA encod
C	35	54	12.8	1089	5	AA875340 DNA encod
C	36	54	12.8	1089	5	AA875103 DNA encod
C	37	54	12.8	1328	5	AA864807 DNA encod
C	38	54	12.8	237961	6	ABQ80552 Human Can
C	39	53.4	12.7	2612	4	AAH17215 Human cDN
C	40	53	12.6	68355	7	ACF62737 Cancer ba
C	41	53	12.6	68355	7	ADB20852 MRP1 base
C	42	53	12.6	68355	9	ADB87941 Human UGT
C	43	53	12.6	68355	9	ADB96924 Human MDR
C	44	53	12.6	68355	9	ADB92115 Human MDR
C	45	53	12.6	186591	7	ACF62750 Cancer ba

ALIGNMENTS

RESULT 1
ACD42206
ID ACD42206 standard; cDNA; 421 BP.
AC ACD42206;
XX
DT 05-SEP-2003 (first entry)
XX
DE Human HLA class I region cDNA Incyte 428822.1.
XX
KW Human; ss; DNA methylation; cancer; colon cancer.
XX
OS Homo sapiens.
XX
FN US2003013099-A1.
XX
PD 16-JAN-2003.
XX
PF 07-MAR-2002; 2002US-00093766.
XX
PR 19-MAR-2001; 2001US-0277380P.
XX
PA (LASEK/) LASEK A K W.
PA (JONE/) JONES D A.
PA (KARP/) KARP A R.
XX
PI Lasek AKW, Jones DA, Karpf AR;
XX
DR WPI; 2003-503249/47.
XX
PT New combination comprising cDNAs that are expressed in a disorder or process associated with DNA methylation, useful for diagnosing, staging, treating or monitoring treatment of cancer, e.g. colon cancer.
XX
PS Claim 2; Page 23; 66pp; English.
XX
CC The invention relates to a combination comprising cDNAs which are expressed in a disorder or process associated with DNA methylation. The combination and cDNAs are useful for diagnosing, staging, treating or monitoring treatment of cancer, e.g. colon cancer and for detecting changes in expression of genes encoding proteins that are associated with DNA methylation. The protein is useful for screening molecules or compounds to identify at least one ligand that binds to the protein and for producing an antibody. The present sequence represents a cDNA expressed in a disorder or process associated with DNA methylation


```

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR P-PSDB; AAO13660.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 1; SEQ ID NO 13651; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and the polypeptides are useful in gene therapy, vaccines or
CC peptide therapies. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 415 BP; 109 A; 99 C; 96 G; 110 T; 0 U; 1 Other;
XX
Query Match 33.5%; Score 141; DB 4; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.6e-34;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 70 ATGCTGTTATGAACCTCCCTTAGAATCAACCCATTGTGAAACACCACTTAATAAGAGC 129
DB 221 ATGCTGTTATGAACCTCCCTTAGAATCAACCCATTGTGAAACACCACTTAATAAGAGC 280
QY 130 AGTATCTTTGGAGATGGAGAGTCACCAAGGATGCCACTGCGTCCCTACAAAGTTTTT 189
DB 281 AGTATCTTTGGAGATGGAGAGTCACCAAGGATGCCACTGCGTCCCTACAAAGTTTTT 340
QY 190 TATGTGAGGACTCGGTCTTCAG 210
DB 341 TATGTGAGGACTCGGTCTTCAG 361
XX
RESULT 4
AAS90405
ID AAS90405 standard; cDNA; 587 BP.
XX
AC AAS90405;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26209.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
FI Drmanac RT, Liu C, Tang YT;
XX

```

```

DR WPI: 2001-639362/73.
DR P-PSDB; ABG26218.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 26209; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological actions in
CC polypeptide and polynucleotide sequences have application of mutations
CC diagnostic, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 587 BP; 152 A; 122 C; 169 G; 144 T; 0 U; 0 Other;
XX
Query Match 13.4%; Score 56.6; DB 5; Length 587;
Best Local Similarity 87.3%; Pred. No. 2.3e-07;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
XX
QY 2 TGGAGTTTCTTCATCTCGTAGGTTCTGTTCTCTCTGCTTCAGGATGAGCTGTAGA 61
DB 476 TGGAGTTTCTTCATCTCGTAGGTTCTGTTCTCTCTGCTTCAGGATGAGCTGTAGA 535
QY 62 ACTCTGCGATG 72
DB 536 CCTTCGTGGTG 546
XX
RESULT 5
AAH18588
ID AAH18588 standard; cDNA; 2566 BP.
XX
AC AAH18588;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18776.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
FN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX
XX 27-AUG-1999; 99JP-00300253.
XX
XX 11-JAN-2000; 2000JP-00118776.
XX
XX 02-MAY-2000; 2000JP-00183767.
XX
XX 09-JUN-2000; 2000JP-00241899.
XX

```

PA (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
PT
PT
XX
XX
PS Claim 8; SEQ ID NO 18776; 2537bp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 2566 BP; 512 A; 568 C; 843 G; 643 T; 0 U; 0 Other;
Query Match 13.4%; Score 56.6; DB 4; Length 2566;
Best Local Similarity 87.3%; Pred. No. 4.3e-07;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGGAGTTCTTCATCTGGTAGGTTCTGCTCTCTCTGCTTCAGGATGAAGCTGTAGA 61
Db 1124 TGGAGTTCTTCATCTGGTAGGTTCTGCTCTCTCTGCTTCAGGATGAAGCTGTAGA 1183
QY 62 ACTCTGCGATG 72
Db 1184 CCTTGGCAGTG 1194
RESULT 6
AAS66944
ID AAS66944 standard; cDNA; 339 BP.
XX
XX AAS66944;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #2748.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-03649167.
XX
XX (HYSE-) HYSERQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABO2757.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
PT
PT
PT
XX
XX Claim 1; SEQ ID NO 2748; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have application in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 339 BP; 45 A; 94 C; 120 G; 80 T; 0 U; 0 Other;
Query Match 13.4%; Score 56.4; DB 5; Length 339;
Best Local Similarity 85.1%; Pred. No. 2.1e-07;
Matches 63; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 TGGAGTTCTTCATCTGGTAGGTTCTGCTCTCTCTGCTTCAGGATGAAGCTGTAGA 61
Db 86 TGGAGTTCTTCATCTGGTAGGTTCTGCTCTCTCTGCTTCAGGATGAAGCTGTAGA 145
QY 62 ACTCTGCGATGCTG 75
Db 146 CCTTGGCAGTCTAG 159
RESULT 7
AAS78088
ID AAS78088 standard; cDNA; 648 BP.
XX
XX AAS78088;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #13892.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX

PN WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Dmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR P-PSDB; ABG00074.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID NO 65; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 648 BP; 126 A; 171 C; 212 G; 139 T; 0 U; 0 Other;
Query Match 13.4%; Score 56.4; DB 5; Length 648;
Best Local Similarity 80.5%; Pred. No. 2.7e-07;
Matches 66; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 3 GGAGTTTCTTCATTTCTGTTAGTTCGTGCTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
DB 87 GGAGTTTCTTCATTTCTGTTAGTTCGTGCTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 146
OY 63 CTTCTCGATGCTGTTATGAAGT 84
DB 147 CTTCTCGATGCTGTTATGAAGT 168
RESULT 10
AAC05498/c
ID AAC05498 standard; cDNA; 380 BP.
XX AAC05498;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 9573.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX

OS Homo sapiens.
XX EPI033401-A2.
PN 06-SEP-2000.
PD 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
PT Claim 1; SEQ ID NO 9573; 71pp + Sequence Listing; English.
PS The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors
XX Sequence 380 BP; 107 A; 107 C; 80 G; 84 T; 0 U; 2 Other;
SQ Query Match 13.1%; Score 55; DB 3; Length 380;
Best Local Similarity 85.9%; Pred. No. 6.1e-07;
Matches 61; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
OY 2 TGGAGTTTCTTCATTTCTGTTAGTTCGTGCTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 61
DB 182 TGGAGTTTCTTCATTTCTGTTAGTTCGTGCTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 123
OY 62 ACTCTGCCGATG 72
DB 122 CCTTCCCGCTG 112
RESULT 11
AAS89076
ID AAS89076 standard; cDNA; 1173 BP.
XX AAS89076;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #24880.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX

PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSB-) HYSEQ INC.
 XX Dmanac RT, Liu C, Tang YT;
 PR PI
 XX
 XX
 XX WPI; 2001-639362/73.
 DR P-PSDB; ABG24889.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX Claim 1; SEQ ID NO 24880; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence novel for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 1173 BP; 204 A; 342 C; 376 G; 251 T; 0 U; 0 Other;
 SQ
 Query Match 13.1%; Score 55; DB 5; Length 1173;
 Best Local Similarity 85.9%; Pred. No. 9.8e-07;
 Matches 61; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 2 TGGAGTTTCTTCTCTGTTAGGTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 61
 DB 170 TGGAGTTTCTTCTCTGTTAGGTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 61
 QY 62 ACTCTCGGATG 72
 DB 230 CCTTCGGGTG 240
 RESULT 12
 AAI19816/c
 ID AAI19816 standard; DNA; 465 BP.
 XX
 XX AAI19816;
 XX
 XX 12-OCT-2001 (first entry)
 XX
 XX Probe #9749 for gene expression analysis in human cervical cell sample.
 DE Probe; human; microarray; gene expression; cervical epithelial cell;
 XX cervical cancer; ss.
 XX Homo sapiens.
 OS
 XX WO200157278-A2.
 PN
 XX
 XX 09-AUG-2001.
 XX

PR 30-JAN-2001; 2001WO-US000670.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human cervical epithelial cells.
 PT
 XX Claim 25; SEQ ID NO 9749; 487pp; English.
 XX
 XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging of
 CC diseases of the cervix, notably cervical cancer. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;
 SQ
 Query Match 13.0%; Score 54.8; DB 4; Length 465;
 Best Local Similarity 79.3%; Pred. No. 7.6e-07;
 Matches 65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 3 GGAGTTTCTTCTCTGTTAGGTCGGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 62
 DB 403 GGAGTTTCTTCTCTGTTAGGTCGGTCTCTGCTTCAGGAATGAAGCTGTAGA 62
 QY 63 CTCTCGGATGCTGTATCAACT 84
 DB 343 CTCTCGGATGCTGTATCAACT 322
 RESULT 13
 AAI10563/c
 ID AAI10563 standard; DNA; 465 BP.
 XX
 XX AAI10563;
 XX
 XX 12-OCT-2001 (first entry)
 XX
 XX Probe #496 for gene expression analysis in human cervical cell sample.
 DE Probe; human; microarray; gene expression; cervical epithelial cell;
 XX cervical cancer; ss.
 XX Homo sapiens.
 OS
 XX WO200157279-A2.
 PN
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000670.
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX

CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 465 BP; 121 A; 129 C; 114 G; 101 T; 0 U; 0 Other;
 Query Match 13.0%; Score 54.8; DB 4; Length 465;
 Best Local Similarity 79.3%; Pred. No. 7.6e-07;
 Matches 65; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 3 GGAGTTTCTTCATTCTGCTAGTTCGTGCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
 DB 403 GGAGTTTCTTCCTTCGTGGTTCGTGCTCTGCTTCAGGAATGAAGCTGTAGAA 62
 QY 63 CTCTGGATGCTGTATGAAC 84
 DB 343 CTTTGGGTGAATGTACAACT 322

Search completed: April 25, 2004, 08:46:06
 Job time : 114.862 secs

GenCÔre version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 1108.58 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-16

Perfect score:

Sequence: 1 atggagtttcttcattctgg.....ttctatgtccttccagcgag 421

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq	length: 0
Maximum DB seq	length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
```

```

29: em_htg_hum:*
30: em_htg_inv:*
31: em_htg_other:*
32: em_htg_mus:*
33: em_htg_pln:*
34: em_htg_rod:*
35: em_htg_man:*
36: em_htg_vrt:*
37: em_sy:*
38: em_htgo_hum:*
39: em_htgo_mus:*
40: em_htgo_other:*
41:

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
C 1	217.8	51.7	193829	9	AC0121534	AC012154	Hot
C 2	217.8	51.7	218508	9	AC020639	AC020639	Hot
C 3	59.6	14.2	163497	9	AC011829	AC011829	Hot
C 4	58.6	13.9	78031	9	AL1392089	AL1392089	Hot
C 5	58.2	13.8	143687	9	AC016572	AC016572	Hot
C 6	58.2	13.8	146078	9	AC091874	AC091874	Hot
C 7	57.8	13.7	174994	2	AC025306	AC025306	Hot
C 8	57.2	13.6	145390	2	AC092963	AC092963	Hot
C 9	57.2	13.6	158983	2	AC020598	AC020598	Hot
C 10	57.2	13.6	178975	2	AC023222	AC023222	Hot
C 11	57.2	13.6	179206	9	CNS01D86	AL121556	BAL
C 12	57.2	13.6	204158	9	AL1333383	AL1333383	Hot
C 13	57	13.5	75609	9	AC006456	AC006456	Hot
C 14	57	13.5	172464	2	AL140899	AL140899	Hot
C 15	57	13.5	174895	2	AC147084	AC147084	Pat
C 16	56.6	13.4	2566	6	AX883871	AX883871	Seq
C 17	56.6	13.4	2566	6	BD160580	BD160580	Pr
C 18	56.6	13.4	2566	6	AK024309	AK024309	Hot
C 19	56.6	13.4	3687	9	BC041827	BC041827	Hot
C 20	56.6	13.4	45303	9	AC0923200	AC0923200	Hot
C 21	56.6	13.4	67605	9	AL441927	AL441927	Hot
C 22	56.6	13.4	993030	9	AC007038	AC007038	Hot
C 23	56.6	13.4	104623	9	AC108064	AC108064	Hot
C 24	56.6	13.4	110000	2	BX546444	08	
C 25	56.6	13.4	110227	9	AC121160	AC121160	Continuation
C 26	56.6	13.4	110227	9	AC125287	AC125287	Hot
C 27	56.6	13.4	115168	2	AC084283	AC084283	Hot
C 28	56.6	13.4	126619	2	AC092458	AC092458	Hot
C 29	56.6	13.4	128000	2	AC004046	AC004046	Hot
C 30	56.6	13.4	130726	9	AL365265	AL365265	Hot
C 31	56.6	13.4	130979	9	AL390719	AL390719	Hot
C 32	56.6	13.4	131060	9	HSD9328	AL132671	Hot
C 33	56.6	13.4	147180	2	AC136469	AC136469	Hot
C 34	56.6	13.4	149628	2	AC034128	AC034128	Hot
C 35	56.6	13.4	149730	9	AL157387	AL157387	Hot
C 36	56.6	13.4	156667	9	AL136094	AL136094	Hot
C 37	56.6	13.4	156942	2	AC136470	AC136470	Hot
C 38	56.6	13.4	157099	9	AC002451	AC002451	Hot
C 39	56.6	13.4	166618	9	AC010323	AC010323	Hot
C 40	56.6	13.4	174914	9	AC086874	AC086874	Hot
C 41	56.6	13.4	181026	2	AL591642	AL591642	Hot
C 42	56.6	13.4	181273	9	AC093898	AC093898	Hot
C 43	56.6	13.4	197148	9	AC087651	AC087651	Hot
C 44	56.6	13.4	197352	2	AC024725	AC024725	Hot
C 45	56.6	13.4	210779	9	AC007450	AC007450	Hot

ALIGNMENT'S

RESULT 1

REC'D I
AC012154/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE
ORIGINATOR

ORGANISM

REFERENCE

REFERENCE AUTHORS

AUTHORS

Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Brieval, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Choudhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Garroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuwona, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pull, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Roife, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshkari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, F.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 193829)
Worley, K.C.

Direct Submission
Submitted (21-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 193829)
Worley, K.C.

Direct Submission
Submitted (07-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 193829)
Worley, K.C.

Direct Submission
Submitted (10-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 193829)
Worley, K.C.

Direct Submission
Submitted (28-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 193829)
Worley, K.C.

Direct Submission
Submitted (28-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 30, 2001 this sequence version replaced gi:14547736.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT:

Location/Qualifiers
1. 193829
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-48H24"

misc_feature
1. 2008
/note="overlaps bases 152829..154828 of AC069070"
/function="clone overlap"
244..365
repeat_region
/rpt_family="AluJb"
366..412
repeat_region
/rpt_family="(CA)n"
413..582
repeat_region
/rpt_family="AluJb"
583..584
repeat_region
/rpt_family="(TAA)n"
607..787
repeat_region
/rpt_family="MER20"
complement(788..862)
repeat_region
/rpt_family="MER2"
complement(901..1072)
repeat_region
/rpt_family="MER2"
1684..1896
repeat_region
/rpt_family="HAL1"
complement(1897..2204)
repeat_region
/rpt_family="AluSg1"
2205..2228
repeat_region
/rpt_family="HAL1"
2229..2251
repeat_region
/rpt_family="(CA)n"
2252..2345
repeat_region
/rpt_family="HAL1"
2521..2646
repeat_region
/rpt_family="MLT11"
3005..3126
repeat_region

```
/rpt_family="HAL1"
repeat_region 3181..3228
/rpt_family="TATG"n"
repeat_region 3230..3273
/rpt_family="AT-rich"
repeat_region 3503..3628
/rpt_family="LFR12"
repeat_region 3629..4072
/rpt_family="LFR12"
repeat_region 4600..4699
/standard_names="76067"
complement(7298..7577)
/rpt_family="AluJo"
repeat_region 8422..8458
/rpt_family="AT-rich"
complement(8696..8873)
/rpt_family="MLT1d"
repeat_region 9120..9187
/rpt_family="CATATA"n"
repeat_region 9695..10135
/rpt_family="MLT1C"
repeat_region 10323..10355
/rpt_family="(TG)n"
repeat_region 10355..10398
/rpt_family="(GA)n"
complement(10948..11063)
/rpt_family="LIP4"
repeat_region 11038..11356
/rpt_family="LIPAL2"
repeat_region 12326..12366
/rpt_family="Harlequin"
repeat_region 13048..13086
/rpt_family="GA-rich"
repeat_region 13117..13360
/rpt_family="GA-rich"
repeat_region 14717..14744
/rpt_family="(TA)n"
repeat_region 14745..14778
/rpt_family="MADE1"
repeat_region 14779..14822
/rpt_family="(TA)n"

Query Match 51.7%; Score 217.8; DB 9; Length 193829;
Best Local Similarity 94.9%; Pred. No. 2.2e-54;
Matches 225; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 185 TTTTATCTGAGGACTCGGCTCTCAGATATGCTACTGGATTCTGTTCAAACTGAGTGA 244
Db 33923 TTATTTTATGCTAAATTTTCCAGATATGCTACTGGATTCTGTTCAAACTGAGTGA 33864

QY 245 AGACATCATCTTTCTTGTAGATCTTCCTTGCTTTCTGCTTCTGCGCAAGAGTTATTCCTTG 304
Db 33863 AGACATCATCTTTCTTGTAGATCTTCCTTGCTTTCTGCTTCTGCGCAAGAGTTATTCCTTG 33804

QY 305 ACCCGACTTCCAAAGAGTACTAGTCTCTTTTAGCAGTGAATTAGAACTGATCTGAGTA 364
Db 33803 ACCCGACTTCCAAAGAGTACTAGTCTCTTTTAGCAGTGAATTAGAACTGATCTGAGTA 33744

QY 365 CTACTGTGTGCTGTGTGACTAGTGTGTTTATTTCTTTCTATGTCCTTTCAGCGAG 421
Db 33743 CTACTGTGTGCTGTGTGACTAGTGTGTTTATTTCTTTCTATGTCCTTTCAGCGAG 33687

RESULT 2
AC020639/c
LOCUS Homo sapiens chromosome 3 clone RP11-32L10, WORKING DRAFT SEQUENCE,
DEFINITION 19 unordered pieces.
ACCESSION AC020639
VERSION AC020639.18 GI:14547756
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 218508)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Banks, T., Barbara, J.,
Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buay, C.,
Burch, P., Burrett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshkari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, F., Teifrod, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 218508)
Worley, K.C.
Direct Submission
Submitted (07-JAN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:14333746.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMTW
Center clone name: RP11-32L10
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 9% of reads
Chemistry: Dye-terminator Big Dye; 91% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 230545 bases at least Q40
Consensus quality: 245133 bases at least Q30
Consensus quality: 253836 bases at least Q20
Estimated insert size: 234517; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 4.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
```

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 19 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1	70229:	contig of 70229 bp in length
* 70230	70329:	gap of unknown length
* 70330	103341:	contig of 33012 bp in length
* 103342	103441:	gap of unknown length
* 103442	136164:	contig of 32723 bp in length
* 136165	136264:	gap of unknown length
* 136265	155333:	contig of 19069 bp in length
* 155334	155433:	gap of unknown length
* 155434	160564:	contig of 5131 bp in length
* 160565	160664:	gap of unknown length
* 160665	167685:	contig of 7021 bp in length
* 167686	167785:	gap of unknown length
* 167786	174526:	contig of 6741 bp in length
* 174527	174626:	gap of unknown length
* 174627	178013:	contig of 3387 bp in length
* 178014	178113:	gap of unknown length
* 178114	183160:	contig of 5047 bp in length
* 183161	183260:	gap of unknown length
* 183261	186615:	contig of 3355 bp in length
* 186616	186715:	gap of unknown length
* 186716	190516:	contig of 3801 bp in length
* 190517	190616:	gap of unknown length
* 190617	194999:	contig of 4383 bp in length
* 195000	195099:	gap of unknown length
* 195100	199336:	contig of 4237 bp in length
* 199337	199436:	gap of unknown length
* 199437	203453:	contig of 4017 bp in length
* 203454	203553:	gap of unknown length
* 203554	206727:	contig of 3174 bp in length
* 206728	206827:	gap of unknown length
* 206828	210654:	contig of 3827 bp in length
* 210655	210754:	gap of unknown length
* 210755	213431:	contig of 2677 bp in length
* 213432	213531:	gap of unknown length
* 213532	215931:	contig of 2400 bp in length
* 215932	216031:	gap of unknown length
* 216032	218508:	contig of 2477 bp in length.
FEATURES		
source		
1..218508		
/organism="Homo sapiens"		
/mol_type="genomic DNA"		
/db_xref="taxon:9606"		
/chromosome="3"		
/clones="RP11-32L10"		
ORIGIN		
Query Match	51.7%;	Score 217.8; DB 2; Length 218508;
Best Local Similarity	94.9%;	Pred. No. 2.2e-54;
Matches 225;	Conservative 0;	Mismatches 12; Indels 0; Gaps 0;
QY	185	TTTTTTATGTGAGACTCGGTCTCAGATATGGCTACTGGGATCTCTTCAAACCTGAGTGA 244
Db	165489	TTATTTTATGCTAAATTTTTCAGATATGGCTACTGGGATCTCTTCAAACCTGAGTGA 165430
QY	245	AGACATCATCTCTTTCTTTGAGATCTCTCTGCTTTCTGCGCAAGAGTTATTCCTG 304
Db	165429	AGACATCATCTCTTTCTTTGAGATCTCTCTGCTTTCTGCGCAAGAGTTATTCCTG 165370
QY	305	ACCGGACTTTCCAAAGAGTACTAGTCTCTTTAGCAGTATTAGAACCTGACTGAGTA 364
Db	165369	ACCGGACTTTCCAAAGAGTACTAGTCTCTTTAGCAGTATTAGAACCTGACTGAGTA 165310
QY	365	CTACTGTGTGTGCTTGTGACTAGTGTGGTTTATTTGTTTCTATGTTCTCTTCAGCGAG 421

Db 165309 CTACTGTGTGCTTGTGACTAGTGTGTTTATTTCTATGTTCTCTTCACGGAG 165253

RESULT 3
AC021829/c

LOCUS AC021829 163497 bp DNA linear PRI 23-JAN-2002
DEFINITION Homo sapiens chromosome 11, clone RP11-436H16, complete sequence.
ACCESSION AC021829
VERSION AC021829.8 GI:17647005
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 163497)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-436H16
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 163497)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczky, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 163497)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Dec 13, 2001 this sequence version replaced g1:17426390.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www-seq.wi.mit.edu>


```

RESULT 4
AL392089      78031 bp      DNA      linear      PRI 21-DEC-2000
LOCUS
DEFINITION
ACCESSION
AL392089
VERSION
AL392089.7  GI:11967571
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 78031)
Direct Submission
Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Dec 22, 2000 this sequence version replaced gi:11691527.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-31F19 is from the library RPCI-11.1 constructed by the group
of Peter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACs.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-31F19 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-130C19 is at 77932 in this
sequence. The true right end of clone RP11-165F24 is at 100 in this
sequence.
FEATURES
             location/Qualifiers
source      1..78031
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="9"
             /map="P24.1-24.3"
             /clone="RP11-31F19"
             /clone_lib="RPCI-11.1"

ORIGIN
Query Match      13.9%; Score 58.6; DB 9; Length 78031;
Best Local Similarity 82.7%; Pred. No. 1.6e-06;
Matches 67; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      4  GAGTTCTTCATCTGGTAGGTTCTGCTCTCTGCTTCAGGAATGAAGCTGTAGAAC 63
DB      32881 GAGTTCTTCATCTGGTAGGTTCTGCTCTCTGCTTCAGGAATGAAGCTGTAGAAC 63
QY      64  TCTCGCATCTGTTATGAACT 84
DB      32941 TTCGCGGTGAGTCAACGCT 32961

```

```

RESULT 5
AC016572      143687 bp      DNA      linear      PRI 17-OCT-2001
LOCUS
DEFINITION
ACCESSION
AC016572
VERSION
AC016572.6  GI:16195210
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143687)
Direct Submission
Unpublished
2 (bases 1 to 143687)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 143687)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 143687)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Oct 17, 2001 this sequence version replaced gi:8886979.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.2.
STS Content:
WI-15997 G21714.
             location/Qualifiers
source      1..143687
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="5"
             /clone="CTB-22L19"

ORIGIN
Query Match      13.8%; Score 58.2; DB 9; Length 143687;
Best Local Similarity 88.7%; Pred. No. 2e-06;
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      2  TGGAGTTCTTCATCTGCTAGGTTCTGCTCTCTGCTTCAGGAATGAAGCTGTAGA 61
DB      133912 TGGAGTTCTTCCTCTCTGCTGGTTCTGCTCTCTGCTTCAGGAGTGAAGCTGCAGA 133971
QY      62  ACTCTGCGATG 72
DB      133972 CCTTAGAGATG 133982

```

```

RESULT 6
AC091874/c    146078 bp      DNA      linear      PRI 03-APR-2003
LOCUS
DEFINITION
ACCESSION
AC091874
VERSION
AC091874.4  GI:29501850
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146078)
REFERENCE

```

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 146078)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 146078)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 146078)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (03-APR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Apr 3, 2003 this sequence version replaced gi:29150342.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 www-shgc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated total Number of Errors is 0.
 NOTE: Transposon sequencing failed to verify the number of repeat copies 12696-13210. Unsure number of repeat copies 12696-13210.
 FEATURES
 Location/Qualifiers
 1. 146078
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2593A12"
 12696..13210
 /notes="NOTE: Transposon sequencing failed to verify the number of repeat copies 12696-13210. Unsure number of repeat copies 12696-13210."
 ORIGIN
 Query Match 13.8%; Score 58.2; DB 9; Length 146078;
 Best Local Similarity 88.7%; Pred. No. 2e-06;
 Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 2 TGGAGTTCTCTCTGCTAGTCTGCTGCTCTCTGCTTCCAGGATGAGCTGTAGA 61
 Db 90111 TGGAGTTCTCTCTCTGCTAGTCTGCTGCTCTCTGCTTCCAGGATGAGCTGTAGA 90052
 QY 62 ACTCTGCGATG 72
 Db 90051 CCTTAGAGATG 90041
 RESULT 7
 AC025306/c 174994 bp DNA linear HTG 26-MAY-2000
 LOCUS AC025306 174994 bp DNA linear HTG 26-MAY-2000
 DEFINITION Homo sapiens clone RP11-24F7, WORKING DRAFT SEQUENCE, 20 unordered pieces.
 ACCESSION AC025306
 VERSION AC025306.3 GI:8072565
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 174994)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 TITLE Homo sapiens, clone RP11-24F7
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 174994)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Menesius,J., Mihova,T., Miranda,C., Mianga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Theodore,J., Tirtell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7328744.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4449
 Center clone name: 24.F.7
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 164074 bases at least Q40
 Consensus quality: 169122 bases at least Q30
 Consensus quality: 171310 bases at least Q20
 Insert size: 172000; agarose-fp
 Insert size: 173094; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 2060: contig of 2060 bp in length
 * 2061 2160: gap of 100 bp
 * 2161 3367: contig of 1207 bp in length
 * 3368 3467: gap of 100 bp
 * 3468 5552: contig of 2085 bp in length
 * 5553 5652: gap of 100 bp
 * 5653 7724: contig of 2072 bp in length
 * 7725 7824: gap of 100 bp
 * 7825 11645: contig of 3821 bp in length
 * 11646 11745: gap of 100 bp
 * 11746 16959: contig of 5214 bp in length
 * 16960 17059: gap of 100 bp
 * 17060 21538: contig of 4479 bp in length
 * 21539 21638: gap of 100 bp
 * 21639 27069: contig of 5431 bp in length
 * 27070 27169: gap of 100 bp
 * 27170 32103: contig of 5934 bp in length
 * 32104 33203: gap of 100 bp
 * 33204 39795: contig of 6592 bp in length
 * 39796 39895: gap of 100 bp
 * *


```

* 39896 4583: contig of 6888 bp in length
* 46584 4683: gap of 100 bp
* 46684 55012: contig of 8329 bp in length
* 55013 55112: gap of 100 bp
* 55113 63297: contig of 8185 bp in length
* 63298 63397: gap of 100 bp
* 63398 74634: contig of 11237 bp in length
* 74635 74734: gap of 100 bp
* 74735 84510: contig of 9776 bp in length
* 84511 84610: gap of 100 bp
* 84611 95119: contig of 10509 bp in length
* 95120 95220: gap of 100 bp
* 95220 106937: contig of 11718 bp in length
* 106938 107037: gap of 100 bp
* 107038 123446: contig of 15309 bp in length
* 12347 123446: gap of 100 bp
* 12347 145374: contig of 22928 bp in length
* 145375 145474: gap of 100 bp
* 145475 174994: contig of 29520 bp in length.
FEATURES
    source
        1. 174994
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="RP11-247"
            /clone_lib="RPC1-11 Human Male BAC"
        1. 2060
            /note="assembly_fragment"
        2161. 3367
            /note="assembly_fragment"
        /note="assembly_fragment"
        clone_end:17
        vector_side:right
        3468. 5552
            /note="assembly_fragment"
        5653. 7724
            /note="assembly_fragment"
        7825. 11645
            /note="assembly_fragment"
        11746. 16959
            /note="assembly_fragment"
        17060. 21538
            /note="assembly_fragment"
        21639. 27069
            /note="assembly_fragment"
        27170. 33103
            /note="assembly_fragment"
        33204. 39795
            /note="assembly_fragment"
        39896. 46583
            /note="assembly_fragment"
        46684. 55012
            /note="assembly_fragment"
        55113. 63297
            /note="assembly_fragment"
        63398. 74634
            /note="assembly_fragment"
        74735. 84510
            /note="assembly_fragment"
        clone_end:SP6
        vector_side:left
        84611. 95119
            /note="assembly_fragment"
        95220. 106937
            /note="assembly_fragment"
        107038. 122346
            /note="assembly_fragment"
        122447. 145374
            /note="assembly_fragment"
        145475. 174994
            /note="assembly_fragment"

```

ORIGIN

Query Match 13.7%; Score 57.8; DB 2; Length 174994;

```

Best Local Similarity 89.9%; Pred. No. 2.7e-06;
Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 GAGTTCTTCATTCTGGTAGGTTCTGCTCTCTCTGCTCAGGATCAAGCTGTAGAAC 63
DB 105787 GAGTTCTTCATTCTGGTAGGTTCTGCTCTCTCTGCTCAGGATCAAGCTGTAGAAC 105728

QY 64 TCTGCGATG 72
DB 105727 TTCGGGGTG 105719

RESULT 8
AC092963 145390 bp DNA linear PRI 30-SEP-2002
LOCUS Homo sapiens 3q BAC RP11-396A13 (Roswell Park Cancer Institute
DEFINITION Human BAC Library) complete sequence.
ACCESSION AC092963
VERSION AC092963.4 GI:18958598
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145390)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,I.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M.,
Louissegue,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,B., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,I., Rivers,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E.,
Sonaika,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamarisa,A., Tameria,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 145390)
Morley,K.C.
Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```



```

QY 3 GGAGTTTCTTCATCTCTGTTAGTGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
Db 20401 GGAGTTTCTTCATCTCTGTTAGTGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
QY 63 CTCGCGATG 72
Db 20461 CTCGCGATG 20470

RESULT 9
AC020598 158983 bp DNA linear HTG 18-FEB-2001
LOCUS Homo sapiens chromosome 2 clone RP11-404J16, WORKING DRAFT
DEFINITION SEQUENCE, 12 unordered pieces.
ACCESSION AC020598
VERSION AC020598.6 GI:12963047
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 158983)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 18, 2001 this sequence version replaced gi:9858987.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0404J16
----- Summary Statistics -----
Sequencing vector: M13; 85%
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 85% of reads
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 153629 bases at least Q40
Consensus quality: 155139 bases at least Q30
Consensus quality: 155960 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 157893; sum-of-contigs
Quality coverage: 6.88 in Q20 bases; agarose-fp
Quality coverage: 6.34 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2675: contig of 2675 bp in length
* 2676 2775: gap of unknown length
* 2776 7991: contig of 5216 bp in length
* 7992 8091: gap of unknown length
* 8092 14234: contig of 6143 bp in length
* 14235 14335: gap of unknown length
* 14336 28444: contig of 14110 bp in length
* 28445 28545: gap of unknown length
* 28546 49191: contig of 20647 bp in length
* 49192 49292: gap of unknown length
* 49293 69715: contig of 20424 bp in length
* 69716 69815: gap of unknown length

Query Match 13.6%; Score 57.2; DB 2; Length 158983;
Best Local Similarity 88.6%; Pred. No. 4e-06; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 8;

QY 3 GGAGTTTCTTCATCTCTGTTAGTGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
Db 82864 GGAGTTTCTTCATCTCTGTTAGTGGTCTCTCTGCTTCAGGAATGAAGCTGTAGAA 62
QY 63 CTCGCGATG 72
Db 82924 CTCGCGATG 82933

RESULT 10
AC023222 178975 bp DNA linear HTG 17-MAY-2000
LOCUS Homo sapiens chromosome 7 clone RP11-396A13 map 7, WORKING DRAFT
DEFINITION SEQUENCE, 12 unordered pieces.
ACCESSION AC023222
VERSION AC023222.3 GI:7884386
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 178975)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 7, clone RP11-396A13
Unpublished
2 (bases 1 to 178975)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dekar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 17, 2000 this sequence version replaced 91:7622394.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6676

Center clone name: 396A.13

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 171366 bases at least Q40

Consensus quality: 175529 bases at least Q30

Consensus quality: 176962 bases at least Q20

Insert size: 192000; agarose-fp

Insert size: 177875; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1862: contig of 1862 bp in length
* 1863: gap of 100 bp
* 1963 5596: contig of 3634 bp in length
* 5597 5696: gap of 100 bp
* 5697 12695: contig of 6999 bp in length
* 12696 12795: gap of 100 bp
* 12796 24148: contig of 11353 bp in length
* 24149 24248: gap of 100 bp
* 24249 34426: contig of 10178 bp in length
* 34427 34526: gap of 100 bp
* 34527 44838: contig of 10312 bp in length
* 44839 44938: gap of 100 bp
* 44939 58360: contig of 13442 bp in length
* 58361 58481: gap of 100 bp
* 58481 79218: contig of 20738 bp in length
* 79219 79318: gap of 100 bp

* 79319 97949: contig of 18631 bp in length
* 97950 98049: gap of 100 bp
* 98050 117961: contig of 19912 bp in length
* 117962 118061: gap of 100 bp
* 118062 144166: contig of 26105 bp in length
* 144167 144267: gap of 100 bp
* 144267 178975: contig of 34709 bp in length.
FEATURES
source
1. 178975
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-396A13"
/clone_lib="RP11-396A13 Human Male BAC"
1. 1862
/note="assembly_fragment"
misc_feature
1963. 5596
/note="assembly_fragment"
misc_feature
5697. 12695
/note="assembly_fragment"
misc_feature
12796. 24148
/note="assembly_fragment"
misc_feature
24249. 34426
/note="assembly_fragment"
misc_feature
24249. 34426
/note="assembly_fragment"
misc_feature
34527. 44838
/note="assembly_fragment"
misc_feature
44939. 58380
/note="assembly_fragment"
misc_feature
58481. 79218
/note="assembly_fragment"
misc_feature
79319. 97949
/note="assembly_fragment"
misc_feature
98050. 117961
/note="assembly_fragment"
misc_feature
118062. 144166
/note="assembly_fragment"
misc_feature
144267. 178975
/note="assembly_fragment"

ORIGIN

Query Match 13.6%; Score 57.2; DB 2; Length 178975;
Best Local Similarity 88.6%; Pred.No. 4e-06; 8; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0; Gaps 0;

Qy 3 GGAGTTTCTTCATCTCGTAGTTCGTGCTCTCTCGGTTTCAGGAATGAAGCTGAGAA 62

Db 6096 GGAGTTTCTTCATCTCGTAGTTCGTGCTCTCTCGGTTTCAGGAATGAAGCTGAGAC 6155

Qy 63 CTCGCGATG 72

Db 6156 CTCGCGATG 6165

RESULT 11

CNS01DS6 179206 bp DNA linear PRI 18-APR-2002
LOCUS BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 367K01
DEFINITION of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION AL121656
VERSION AL121656.2 GI:7159617
KEYWORDS SPG4 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179206)
AUTHORS Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,

Artiguenave, R., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P., Brottier, P., Cattolico, L., Barbe, V., Burgunder, J.M., Prud'Homme, J.F., Brice, A., Fontaine, B., Heilig, R. and Weissenbach, J.

TITLE
Spastin, a novel AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia

JOURNAL
Nat. Genet. (1999) In press

REFERENCE
2 (bases 1 to 179206)

AUTHORS
Genoscope.

TITLE
Direct Submission

JOURNAL
Submitted (18-APR-2002) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
On Mar 6, 2000 this sequence version replaced gi:6981752.

FEATURES
Location/Qualifiers
1..179206
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="367K01"
/clone_lib="CITB_978_SKB"

ORIGIN

Query Match 13.6%; Score 57.2; DB 9; Length 179206;
Best Local Similarity 88.6%; Pred. No. 4e-06;
Matches 62; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

3 GGAGTTTCCTCATCTCGTAGGTCGTGCTCTCTGGCTCAGGATGAGCTGTAGAA 62
Db 107000 GGAGTTTCCTCTTCGTGGTGGTTCGGTCTCAGGCTCAGGAGTGCAGAC 107059
Qy 63 CTTCTGGGATG 72
Db 107060 CTTTGGCGTG 107069

RESULT 12

AL133383
LOCUS
DEFINITION
Human DNA sequence from clone GSI-204112 on chromosome 1. Contains ESTs, STSs and GSSs. Contains a novel gene, complete sequence.

ACCESSION
AL133383

VERSION
AL133383.10 GI:10119658

KEYWORDS
HTG.

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 204158)
Hall, R.

REFERENCE
1 (bases 1 to 204158)

AUTHORS
Direct Submission

TITLE
Submitted (05-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

JOURNAL
requests: clonerequest@sanger.ac.uk

COMMENT
On Sep 12, 2000 this sequence version replaced gi:10086005.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SwissProt; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep GSI-204112 is from the library genome Systems Release VECTOR: pBelobAC11
This sequence is the entire insert of clone GSI-204112 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is

ambiguous, there is an annotation using the 'unsure' feature key. This sequence was generated by the Sanger Centre from part of a human chromosome 1 bacterial clone contig constructed by John Capten et al, NHGRI, NIH. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1.

FEATURES

source
1..204158
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="GSI-204112"
/clone_lib="Genome_Systems_Release1"
7..442
/note="match: GSS: Em:AQ761971"
245..300
/note="14 copies 4 mer tata 73% conserved"
251..298
/note="6 copies 8 mer tatacata 85% conserved"
256..301
/note="23 copies 2 mer at 80% conserved"
313..989
/note="L1ME3A repeat: matches 5406. .6152 of consensus"
1047..1513
/note="match: GSS: Em:AQ819686"
1908..2053
/note="FAM repeat: matches 26. .172 of consensus"
2276..2574
/note="Alusx repeat: matches 9. .303 of consensus"
2720..2870
/note="L2 repeat: matches 2558. .2711 of consensus"
4330..4437
/note="MER94 repeat: matches 10. .121 of consensus"
4758..4793
/note="18 copies 2 mer tt 91% conserved"
5066..5201
/note="MIR repeat: matches 7. .140 of consensus"
6972..7446
/note="L1MEC repeat: matches 278. .780 of consensus"
7453..8545
/note="HERVL repeat: matches 2899. .4004 of consensus"
8794..9044
/note="HERVL repeat: matches 2335. .2589 of consensus"
9613..9666
/note="Alu repeat: matches 1. .53 of consensus"
9667..9904
/note="MLT1D repeat: matches 168. .416 of consensus"
9916..10048
/note="AluJ0 repeat: matches 1. .136 of consensus"
10884..11146
/note="Tigger3(Golem) repeat: matches 2771. .3027 of consensus"
11160..11284
/note="Tigger3(Golem) repeat: matches 1. .125 of consensus"
11348..11672
/note="AluJb repeat: matches 1. .310 of consensus"
11707..11794
/note="MIR repeat: matches 4. .92 of consensus"
11798..12173
/note="L1FA16 repeat: matches 5717. .6155 of consensus"
12188..12287
/note="MIR repeat: matches 138. .235 of consensus"
12471..16388
/note="L1P5 repeat: matches 2293. .6155 of consensus"
16841..16898
/note="L2 repeat: matches 2693. .2750 of consensus"
18264..18563
/note="Alusq repeat: matches 1. .300 of consensus"
18637..19137
/note="L1ME2 repeat: matches 5609. .6131 of consensus"
19147..19358
/note="53 copies 4 mer tata 75% conserved"
19391..19835

TITLE Direct Submission
JOURNAL Submitted (04-JUN-2002) Department of Genetics, Washington
REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS 7 (bases 1 to 75609)
TITLES Waterston, R.
JOURNAL Direct Submission
COMMENT Submitted (26-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 4, 2002 this sequence version replaced gi:4454519.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0969D04

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GrB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is H_DJ0558L10, 200 bp overlap.
Actual start of this clone is at base position 75969 of RP4-558L10
actual end is at base position 75609 of RP5-969D4.

FEATURES

source	1..75609	/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="7"
		/map="7"
		/clone="RP5-969D4"
		/clone_lib="RPCI-4"
repeat_region	601..948	/rpt_family="L1"
variation	921	/allele="G"
		/alleles="A"
		/db_xref="dbSNP:1544590"
repeat_region	1026..1110	/rpt_family="GA-rich"
variation	1046	/allele="A"

repeat_region	1113..1256	/db_xref="dbSNP:1544591"
variation	1264	/rpt_family="A-rich"
		/allele="C"
		/alleles="T"
		/db_xref="dbSNP:2537200"
repeat_region	1507..1528	/rpt_family="AT-rich"
repeat_region	1928..2057	/rpt_family="L1"
repeat_region	3052..3266	/rpt_family="Alu"
variation	3806	/allele="G"
		/alleles="A"
		/db_xref="dbSNP:2727626"
variation	3887	/allele="T"
		/alleles="C"
		/db_xref="dbSNP:2727625"
variation	4632	/allele="A"
		/alleles="T"
		/db_xref="dbSNP:2727624"
variation	4914	/alleles="C"
		/alleles="T"
		/db_xref="dbSNP:957279"
variation	5067	/alleles="C"
		/alleles="T"
		/db_xref="dbSNP:768816"
variation	5080	/alleles="C"
		/alleles="T"
		/db_xref="dbSNP:768817"
repeat_region	5477..5678	/rpt_family="L1"
repeat_region	5679..5973	/rpt_family="Alu"
repeat_region	5974..6098	/rpt_family="L1"
repeat_region	6101..6431	/rpt_family="L1"
repeat_region	6432..6736	/rpt_family="Alu"
variation	6481	/alleles="T"
		/alleles="A"
		/db_xref="dbSNP:726749"
variation	6598	/alleles="G"
		/alleles="A"
		/db_xref="dbSNP:726750"
repeat_region	6737..6845	/rpt_family="L1"
variation	6749	/alleles="T"
		/alleles="C"
		/db_xref="dbSNP:726751"
variation	7861	/alleles="C"
		/alleles="A"
		/db_xref="dbSNP:2727632"
repeat_region	8094..8163	/rpt_family="AcHobo"
repeat_region	8463..8511	/rpt_family="AT-rich"
variation	10046	/allele="A"

```
/allele="G"
/db_xref="dbSNP:2537199"
10459..110758
/rpt_family="Alu"
11081
/allele="G"
/allele="A"
12749..13573
/db_xref="dbSNP:887689"
/rpt_family="ERV1"
13686..13848
/rpt_family="ERV1"
14311..14339
/rpt_family="(TTTTG)n"
14516..14702
/rpt_family="MIR"
16054..16353

Query Match      13.5%; Score 57; DB 9; Length 75609;
Best Local Similarity 86.3%; Pred.No. 4.7e-06;
Matches 63; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGGAGTTCTTCATCTCGTAGGTCGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 61
Db 12878 TGGAGTTCTTCATCTCGTAGGTCGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 12937

QY 62 ACTCTGCGATGCT 74
Db 12938 CCTTGGCGGTGT 12950

RESULT 14
AC140889      172464 bp DNA linear HTG 05-MAR-2003
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-171G19, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION AC140889
VERSION AC140889.1 GI:28850000
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 172464)
DOE Joint Genome Institute.
Direct Submission
Submitted (05-MAR-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 472192
Center clone name: RPCI-11_171G19
-----
Summary Statistics
Consensus quality: 167507 bases at least Q40
Consensus quality: 168805 bases at least Q30
Consensus quality: 169686 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 171564; sum-of-contigs estimation
Quality coverage: 5.88 in Q20 bases; agarose-fp estimation
Quality coverage: 6.0 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1522: contig of 1522 bp in length
* 1523: gap of unknown length
* 1623: contig of 1513 bp in length
* 3135: gap of unknown length
* 3235: contig of 3301 bp in length
* 6535: gap of unknown length
* 6635: gap of unknown length
* 6637: contig of 8479 bp in length
* 13115: gap of unknown length
* 13215: gap of unknown length
* 13216: contig of 10766 bp in length
* 23981: gap of unknown length
* 23982: contig of 8779 bp in length
* 24081: gap of unknown length
* 32860: contig of 8779 bp in length
* 32861: gap of unknown length
* 32951: contig of 9351 bp in length
* 42311: gap of unknown length
* 42312: contig of 25660 bp in length
* 42412: gap of unknown length
* 68072: contig of unknown length
* 68172: contig of 51064 bp in length
* 119235: gap of unknown length
* 119236: contig of unknown length
* 119335: gap of unknown length
* 119336: contig of 53129 bp in length.

FEATURES
Location/Qualifiers
1..172464
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-171G19"
/clone_lib="RPCI human BAC library 11"

ORIGIN
Query Match      13.5%; Score 57; DB 2; Length 172464;
Best Local Similarity 86.3%; Pred.No. 4.6e-06;
Matches 63; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 TGGAGTTCTTCATCTCGTAGGTCGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 61
Db 21559 TGGAGTTCTTCATCTCGTAGGTCGTCTCTCTGCTTCAGGAATGAAGCTGTAGA 21618

QY 62 ACTCTGCGATGCT 74
Db 21619 CCTTGGCGGTGT 21631

RESULT 15
AC147084/c
LOCUS
DEFINITION Pan troglodytes chromosome 7 clone RP43-64A1, *** SEQUENCING IN
PROGRESS ***, 9 unordered pieces.
ACCESSION AC147084
VERSION AC147084.1 GI:38154122
KEYWORDS HTG; HTGS_PHASE1 (chimpanzee)
SOURCE Pan troglodytes
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Wilson, R.K.
1 (bases 1 to 174895)
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 174895)
Wilson, R.K.
Direct Submission
Submitted (04-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
-----Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
```



```

175579 174895: Contig of 1317 bp in length.
Location/Qualifiers
1..174895
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/clones="RP43-64A1"
1..2971
/note="assembly_name:Contig10"
3072..12999
/note="assembly_name:Contig11"
13100..30461
/note="assembly_name:Contig12"
30562..60139
/note="assembly_name:Contig13"
60240..92522
/note="assembly_name:Contig14"
92623..170864
/note="assembly_name:Contig15"
170965..172086
/note="assembly_name:Contig7"
172187..173478
/note="assembly_name:Contig8"
173579..174895
/note="assembly_name:Contig9"

ORIGIN
Query Match 13.5%; Score 57; DB 2; Length 174895;
Best Local Similarity 86.3%; Pred. No. 4,6e-06;
Matches 63; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 2 TGGAGTTTTCATCTGGTAGGTTCTGTGCTCTCTGCGCTCAGCAATCAGCTGTAGA 61
Ddb 131899 TGGAGTTTTCCTTCCTTCGTGGGTTCTGTGCTCTCTGAGAAATCAAGCTGCAGA 131840
QY 62 ACTCTGCGCATGCT 74

```

	1	615	38.0	772	12	BG431098	602498723
C	2	613.8	37.9	745	13	B0681294	UI-CF-EC1
C	3	600.2	37.1	685	29	AG165287	Pan trogl
C	4	577.6	35.7	696	14	CA312303	UI-CF-FN0
						BG431098	602498723

FEATURES
source

Location/Qualifiers
1. 772
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="IWAGB:4612420"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_75"
/notes="Organ: kidney; Vector: pDNR-LIB (Clontech); Site:1: SfiI (ggccattatggcc); Site:2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGC-3, and 3' adaptor sequence: 5'-ATTCTAGAGCGCAGGCGCCGATG-dt(30)BN-3' (where B = A, C, or G, and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies for full-length clones by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 38.0%; Score 615; DB 12; Length 772;
Best Local Similarity 95.4%; Pred. No.1.2e-114;
Matches 730; Conservative 0; Mismatches 25; Indels 10; Gaps 9;

QY 508 GTACATTAAGTGTGCATCTCTTCCTTCCCTAAATAATGATTTTACTGCTTGAATATACAT 567
DB 1 GTAGATTAAAGTGTGCATTC-TTGTTCCTTAAATATGATTTTACTGCTTGAATATACAT 59

QY 568 TTGAGTTGAAGTTTAGAAACTAAACATAGCATTAATATGATTAATGATGCAATGGAATATTATTA 627
DB 60 TTGAGTTGAAGTTTAGAAACTAAACATAGCATTAATATGATTAATGATGCAATGGAATATTATTA 118

QY 628 TCCCTTTGAAAACTGATTGATAAAATATATTCCTCCCTCCTTTAGAAACAGTCAAAGCCACT 687
DB 119 TCCCTTTGAAACTGATTGATAAAATATATTCCTCCCTCCTTTAGAAACAGTCAAAGCCACT 178

QY 688 TCAACCAAGTTTCAAAATAAAGGAAGGTGACGAGTTAGGGCATGCGATTAATATTTCTTGG 747
DB 179 TCAACCAAGTTTCAAAATAAAGGAAGGTGACGAGTTAGGGCATGCGATTAATATTTCTT-G 237

QY 748 CTTCGTTGATACCCATTGGCCAGGCGCTTTATAAGGACTCCCAAAGCATTTTGAAGAAT 807
DB 238 CTTCGTTGTATA-CCAGTTGTCAAGGACATATATAGGACTCCCAAAGCATTTTGAAGAAT 296

QY 808 GGCATATCAAAATAGTGTATGTCCTCTCAAAATGAGGCATTTTAAATGTTTCAATCTAT 867
DB 297 GGCATATCAAAATAGTGTATGTCCTCTCAAAATGAGGCATTTTAAATGTTTAAATCTAT 356

QY 868 TTGACGCTCAGGTTATGATATGTTATGAAAAATAAGCTTCATATTTCCTTATAGCTAC 927
DB 357 TTGACGCTCAGGTTATGATATGTTATGAAAAATAAGCTTCATATT-TTATAGCTAC 415

QY 928 ATCCTATTATTCCTTTTAGAAAAACAGATTAACAAATAGTTTTTAATAGTTGCCATCTTA 987
DB 416 ATCCTATTATTCCTTTTAGAAAAACAGATTAACAAATAGTTTTTAATAGTTGCCATCTTA 475

QY 988 GCATTTATCAGGTCTAATGAAACCAATATGAAATCTCTGTATAAATATTTTCTGATGTTAC 1047
DB 476 GCATTTATCAGGTCTAATGAAACCAATATGAAATCTCTGTATAAATATTTTCTGATGTTAC 535

QY 1048 TAGCTATGGGAAATTAGNACTGGCACACACCTCTGACATCTACTAAGTGGAAATGTTAGGAT 1107
DB 536 TAGCTATGGGAAATTAGNACTGGCACAA-CTGACATTTACTTAAGTGGAAATGTTAGGAAAT 594

QY 1108 TTTCCGCATCGCATGTTAGAAATCTCTAAAATTTTAAACA-TTCCTGTTTAAATGACTAAGGT 1166
DB 595 TTTCCGCATCGCATGTTAGAAATCTCTAAAATTTTAAACATTTTCCCTGTTTAAATGACTAAGGT 654

QY 1167 TTGCTTTTATCAATATGAAATTTGAGGCCAATATCATACATTAACATATGAAAGCTTTT 1226
DB 655 TTGCTTTTATCAATATGAAATTTCTGAGGCCAATATCATACATTAACATTA--TATGAAAGTTTT 712

QY 1227 AATTCCTAAAAATAGTTTTAGAGATATTTCAAGCAATGCTCTCCTTA 1271

[illegible]

```

2 (bases 1 to 685)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1..685
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-032L08.T7"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN
Query Match 37.1%; Score 600.2; DB 29; Length 685;
Best Local Similarity 96.1%; Pred. No. 1.2e-111;
Matches 647; Conservative 0; Mismatches 23; Indels 3; Gaps 3;
QY 660 CTTCTTTAGAACAGCTCAAAGCCACTTCAAACAAGTTTCAAAATAAGAGGAGGTAGCA 719
DB 684 CTTCTTTAGAACAGCTCAAAGCCACTTCAAACAAGTTTCAAAATAAGAGGAGGTAGCA 625
QY 720 AGTTAGCGCATGATATATTTCTTGCGTTGTTGTATACCCATGCGCAGGCGCTTAT 779
DB 624 AGTTAGCGCATGATATATTTCTT-GCTTGTGTATA-CCAGTTGTCAGGACATTAT 567
QY 780 AAGGACTCCCAAAGCAATTTTGAAGAAATGCAATATCAAAATAAGTGTATGTCCTCAAA 839
DB 566 AAGGACTCCCAAAGCAATTTTGAAGGATGCAATATCAAAATAAGTGTATGTCCTCAAA 507
QY 840 TGAGGCATTTTAAATGCTTCAATCTATTTGGAGCTCAGGTTATGATATGTTTATGAAA 899
DB 506 TGAGGCATTTTAAATGCTTAAACCTATTTGGACCCCTCAGGTTATGATATGTTTATGAAA 447
QY 900 AATAAGCTTCATTATTTCTTTATAGTACATCCTATTATTCCTTTTGAACAACAGATAA 959
DB 446 AATAAGCTTCATTATTT-TTATCGTACATCCTATTATTCCTTTTGAACAACAGATAA 388
QY 960 CAATAAGTTTTATAGTTGTCATATTTAGCATTTATCAGGCTCAATGAACCAATATTGA 1019
DB 387 CAATAAGTTTTATAGTTGTCATATTTAGCATTTATCAGGCTCAATGAACCAATATTGA 328
QY 1020 ATCTCTGATAAATATTTCTGATGTTTACTAGCTATGGGAAATTAGAACTGCAACCCCT 1079
DB 327 ATCTCTGATAAATATTTCTGATGTTTACTAGCTATGGGAAATTAGAACTGCAACCCCT 268
QY 1080 GACATTACTAAGTGGAAATGTTAGGATTTTTCGCATCGCATGTTAGAACTCTAAATTT 1139
DB 267 GACATTACTAAGTGGAAATGTTAGGATTTTTCGCATCGCATGTTAGAACTCTAAATTT 208
QY 1140 TAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAAATTCCTGAAGCCCAAT 1199
DB 207 TAAACATTCCTGTTAAATGACTAAGGTTTGCTTTTATCAATATGAAATTCCTGAAGCCCAAT 148
QY 1200 ATCATACATTAACCTAGAAAGCTTTTAAATTCCTAAATAATAGTTTTAGAGATTTCAAGC 1259
DB 147 ATCATACCTTAACTATGAAGCTTTTAAATTCCTAAATAATAGTTTTAGAGATTTCAAGC 88

```

```

QY      1260 AATGCTCTCTATATATCCATACCAAGTGTGTTATGACACAAATTCACCTAGTCTGGTTT 1319
Db      |||||||
87      AATGCTCTCTCTATATATCCATACCAAGTGTGTTATGACACAAATTCACCTAGTCTGGTTT 28
QY      1320 AAAAAATGAATCT 1332
Db      |||||||
27      AATGGAATCT 15

RESULT 4
CA312303/c
LOCUS      696 bp mRNA linear EST 04-NOV-2002
DEFINITION UI-CF-FNO-afm-n-21-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
            UI-CF-FNO-afm-n-21-0-UI 3', mRNA sequence.
ACCESSION  CA312303
VERSION     CA312303.1 GI:24530401
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 696)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL   Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE   Normalization and subtraction: two approaches to facilitate gene
PUBMED    Genome Res. 6 (9), 791-806 (1996)
COMMENT   Email: paul-mccray@uiowa.edu
           Tissue Procurement: Dr. M. J. Welsh, University of Iowa
           cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
           cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
           Clone Distribution: Researchers may obtain clones from Research
           Genetics (www.resgen.com) or from Open Biosystems
           (www.openbiosystems.com).
           The following repetitive elements were found in this cDNA
           sequence: 1-35, >AT rich#Low_complexity
           Seq primer: M13 FORWARD
           PolyA=yes.

FEATURES             Location/Qualifiers
     1..696
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="UI-CF-FNO-afm-n-21-0-UI"
     /tissue_type="Human Lung Epithelial cells"
     /lab_host="DH10B (Life technologies) (T1 phage resistant)"
     /clone_lib="UI-CF-FNO"
     /note="Organ: Lung; Vector: p77T3-Pac (Pharmacia) with a
           modified polylinker; Site 1: EcoR I; Site 2: Not I;
           UI-CF-FNO is a subtracted cDNA library derived from two
           normalized Human lung epithelial cell libraries (EN1 and
           DUL). The library was subtracted according to according to
           Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
           1996. For additional information, contact:
           bento-soares@uiowa.edu
           TAG TISSUE=Lung Epithelial Cells Tissue nos 359-368
           TAG LIB=UI-CF-FNO
           TAG_SEQ=GGCTGTAGGC"

ORIGIN
Query Match      35.7%; Score 577.6; DB 14; Length 696;
Best Local Similarity 95.6%; Pred. No. 4.7e-107;
Matches 668; Conservative 0; Mismatches 22; Indels 9; Gaps 7;

QY      921 TAGCTACATCCCTATTATTCCTTTTAGAAACAAGATAACAATAAGTTTAAATAGTTGCC 980

```

```

Db      |||||||
596      TAGCTACATCCCTATTATTCCTTTTAGAAACAAGATAACAATAAGTTTAAATAGTTGCC 637
QY      |||||||
981      ATACTAGCAATTTATCAGGTCTAATGAACCAATATTGAATCTCTGATAATATTTCTG 1040
Db      |||||||
636      ATACTAGCAATTTATCAGGTCTAATGAACCAATATTGAATCTCTGATAATATTTCTG 577
QY      |||||||
1041      ATGTTACTAGCTATGGGAAATTAGAACTGGCAACACCCCTGACATTTACTAAGTGGAAATGT 1100
Db      |||||||
576      ATGTTACTAGCTATGGGAAATTAGAACTGGCAACACCCCTGACATTTACTAAGTGGAAATGT 517
QY      |||||||
1101      TAGAATTTTTCGGCATCGCATGTAGAAATCTCTAAATTTTAAACATTTCTCTGTTAAATGAC 1160
Db      |||||||
516      TAGAATTTTTCGGCATCGCATGTAGAAATCTCTAAATTTTAAACATTTCTCTGTTAAATGAC 457
QY      |||||||
1161      TAAGGTTTGTCTTTTATCAATATGAATTTCTGAAGGCCAATATCATACCAATTAACCTATGAAA 1220
Db      |||||||
456      TAAGGTTTGTCTTTTATCAATATGAATTTCTGAAGGCCAATATCATACCAATTAACCTATGAAA 397
QY      |||||||
1221      GCTTTTAATTCCTAAATAGTTTATAGATATTTCAAGCAATGCTCTCTCTAATATCCATA 1280
Db      |||||||
396      GCTTTTAATTCCTAAATAGTTTATAGATATTTCAAGCAATGCTCTCTCTAATATCCATA 337
QY      |||||||
1281      CGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGGTTTAAAAATGAAATCTTTATATTG 1340
Db      |||||||
336      CGCAAGTGTGTTTATGACACAAATTCACCTAGTCT- GTTTAAAAATGAATCTTTATATTG 278
QY      |||||||
1341      ACTGGGTCTCCCAATATTCAGTAATTTCTGTTATGAGGAGCTTGAATAGCAAAATTG 1400
Db      |||||||
277      ACT-GGTGTTCCACATATTTTCAGTAATTTCTGTTATGAGGAGCTTGAATAGCAAAATTG 219
QY      |||||||
1401      CCCACACAGTTTAACTGGATAGATCACTGCTGCTGATCATAAACCACTTGGTACTACACC 1460
Db      |||||||
218      -CCACACAGTTTAACTGGATAGACACCGTACCTGGTGATCATACCACTTGGTACTACACC 160
QY      |||||||
1461      CAGAAACTCAAAATTTGCTTTTCTTCCTGATCAGATATGGGTGCTCTTTTGTAGCTCTAG 1520
Db      |||||||
159      CAGAAACTCAAAATTTGCT- TTTCTCCTGATGAGATAT- GGGTGCTCTTTTGTAGCTCTAG 102
QY      |||||||
1521      GGCCTAGGCTACCCCAAGTGAATGATATATAAGCAAAATGTGTTGTATCCAGAGTCTT 1580
Db      |||||||
101      GCC---TAGGTAACCAAGTGGAGTGATATATATAGC-AAATGTGTTGTATCCAGAGTCTT 46
QY      |||||||
1581      CTTGTCATTTGTAATAAAAAATTTATTTAAAAATTTAAAA 1619
Db      |||||||
45      CTTGTCATTTGTAATAAAAAATTTATTTAAAAAATAAAAA 7

RESULT 5
AW179297/c
LOCUS      PM3-ST0115-100999-001-C07 ST0115 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW179297
ACCESSION  AW179297
VERSION     AW179297.1 GI:6445334
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 623)
JOURNAL   HCGP http://www.ludwig.org.br/ORESTES.
COMMENT    The FAPESP/LICR Human Cancer Genome Project
           Unpublished (1999)
           Contact: Simpson A.J.G.
           Laboratory of Cancer Genetics
           Ludwig Institute for Cancer Research
           Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
           Brazil
           Tel: +55-11-2704922
           Fax: +55-11-2707001
           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPESP/LICR Human Cancer Genome
           Project. This entry can be seen in the following URL

```


661 CTCCTTTAGAACAGTCAAAAGCCACTTCAAAAGTTTCAAATTAAGGAAGTAGCAA 720
203 CTCCTTTAGAACAGTCAAAAGCCACTTCAAAAGTTTCAAATTAAGGAAGTAGCAA 144
721 GTTAGGCGATGATATATTTTCTTGGCTTGTGTATACCCATTGCCAGGGCTTTATA 780
143 GTTAGGCGATGATATATTTTCTT-GCTTGTGTATA-CCAGTGTGCAAGAACATTATA 86
781 AGGACTCCAAAGCAATTTTCAAGAAAGGCAATATCAAATAAGTGTATGTCCTCTCAAT 840
85 AGGACTCCAAAGCAATTTTCAAGAAAGGCAATATCAAATAAGTGTATGTCCTCTCAAT 26
841 GAGGCATTTTAAATGTTACATCT 865
25 GAGGCATTTTAAATGTTAAATCT 1

RESULT 7
LOCUS AI791550
DEFINITION ok82d09.y5 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1520465 5', mRNA sequence.
ACCESSION AI791550
VERSION AI791550.1 GI:5339266
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 595)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Other ESTs: ok82d09.x5
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation information
This 5' resequenced clone has no previous 5' data to verify this new read against
Insert Length: 3866 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 459.
Location/Qualifiers
1: 595
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1520465"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid3"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES
source

ORIGIN
source

Query Match 34.8%; Score 564; DB 9; Length 595;
Best Local Similarity 95.8%; Pred. No. 2.8e-104;
Matches 564; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 23 ATAAATATGCTAATCCATAGAAAAGGATATAAAGTATTAAATATGCTATATAGC 82
DB 1 ATAAATATGCTAATCCATAGAAAAGGATATAAAGTATTAAATATGCTATATAGC 60
QY 83 NNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAAGTCACCTTCAATTAAGAAAACAT 142
DB 61 TATATCTCTATCTATGTATCTAATAGGGAAGTTCAAGTCACCTTCAATTAAGAAAACAT 120
QY 143 ATCTCTGAGCATAGGAGCAGCTCAGTCCCTATGCTGGATGCGAGTGACAGGAGAGGG 202
DB 121 ATCTCTGAGCATAGGAGCAGCTCAGTCCCTATGCTGGATGCGAGTGACAGGAGAGGG 180
QY 203 GAAATTAGAAAAGAGAACTATATATTTGAAAAGGATATAAAGCATTAAATATATGATA 262
DB 181 GAAATTAGAAAAGAGAACTATATATTTGAAAAGGATATAAAGCATTAAATATATGATA 240
QY 263 TATAGCTATATCTATGTATGTATCTTAACAGAGAGTTCAAGTCACCTTCAATTAAGAAAC 322
DB 241 TATAGCTATATCTATGTATGTATCTTAACAGAGAGTTCAAGTCACCTTCAATTAAGAAAC 300
QY 323 ATTTTGTAGCATGGACCCAGCTCAGGTCCTTATGCTGGATGCGAGTAGACAGGAGATGG 382
DB 301 ATTTTGTAGCATGGACCCAGCTCAGGTCCTTATGCTGGATGCGAGTAGACAGGAGATGG 360
QY 383 GAAATTAGAAAAGAGAACTGTGTATTTGAAATGAGTGGCTGACCCCTTAAGGAACCTTA 442
DB 361 GAAATTAGAAAAGAGAACTGTGTATTTGAAATGAGTGGCTGACCCCTTAAGGAACCTTA 420
QY 443 TAATTAATGATCATCTGAAATAAACATACAGGATAAAGATGTCAAATGAGTGTGACTCCC 502
DB 421 TAATTAATGATCATCTGAAATAAACATACAGGATAAAGATGTCAAATGAGTGTGACTCCC 480
QY 503 TTAAGTAGATTAAAGTGTGCAATCTTTGTTTCCCTAAATATGATTTACTGCTTGAAT 562
DB 481 TTAAGTAGATTAAAGTGTGCAATCTTTGTTTCCCTAAATATGATTTACTGCTTGAAT 540
QY 563 TACATTTGAGTTGAAGTTTAGAAATTAACATAGCATTAAATATGAATAAT 611
DB 541 TACATTTGAGTTGAAGTTTAGAAATTAACATAGCATTAAATATGAATAAT 589

RESULT 8
LOCUS BE866341
DEFINITION 601578994F1 NTH_MGC_53 Homo sapiens cDNA clone IMAGE:3961510 5', mRNA sequence.
ACCESSION BE866341
VERSION BE866341.1 GI:10315221
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 717)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM845 row: 1 column: 23
High quality sequence stop: 630.
Location/Qualifiers
1: 717

FEATURES
source

Qy	1495	TATGGG	1500
Db	712	TATGGG	717
RESULT 9			
LOCUS	BC399616		
DEFINITION	602441635F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4557321 5'		
ACCESSION	BC399616		
VERSION	BC399616.1	GI:13292983	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgi.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: InCyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLC1261 row: f column: 10 High quality sequence stop: 575. Location/Qualifiers 1..735 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4557321" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH_MGC_75" /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site:1: SfiI (ggcgcttcgccc); Site 2: SfiI (ggccattagccc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb, range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
FEATURES			
SOURCE			
Query Match	32.8%;	Score 531.8;	DB 12; Length 735;
Best Local Similarity	89.9%;	Pred. No. 9e-98:	
Matches 612;	Conservative	0;	Mismatches 61; Indels 8; Gaps 5;
Qy	1	TTAAATCTACATACTCTTATGTATATAAATATGCTTAATCCATGAAGGATATAAG	60
Db	37	TTAAATCTACATACTCTTATGTATATAAATATGCTTAATCCATGAAGGATATAAG	96
Qy	61	TATTAATATGTCTATATAGCNNNNNNNNNNNNNNNNNNNNNNATAGGAAGTTCAA	120
Db	97	TATTAATATGTCTATATAGCTATATCTATCTATCTATCTATCTATAGGAAGTTCAA	156
Qy	121	GTCACCTTCAATTGAAGAAACATATCTCTGAGCATAGGACGAGCTCCTTAGTGTTG	180
Db	157	GTCACCTTCAATTGAAGAAACATATCTCTGAGCATAGGACGAGCTCCTTAGTGTTG	216
Qy	181	GATGCGAGTCGACAGGAGGGGGAATTAGAAAAGAGACTATATTAATTCGAAAAAGGAT	240
Db	217	GATGCGAGTCGACAGGAGGGGGAATTAGAAAAGAGACTATATTAATTCGAAAAAGGAT	276
Qy	241	ATAAAGCATTAAATATATGATATATAGCTATATCTATGTATGTATCTATCTAACAGAGAAGTTC	300

	Best Local Similarity	Pred.No. 1.8e-88:	Mismatches	Conservative	0;	Gaps	7;
			Matches	574;	Indels	9;	
QY	987	AGGTCCTAATGAACCAATAATTGAATCTCTGTAAATAATTTTCTGATGTTACTAGCTATGG	1056				
Db	589	AGGTCCTAATGAACCAATAATTGAATCTCTGTAAATAATTTTCTGATGTTACTAGCTATGG	530				
QY	1057	GAATATTAGAATCGSCAACCCCTGCACATTACTAAAGTGGAAAATGTTAGGATTTTTCGGCAT	1116				
Db	529	GAATATTAGAATCGSCAACCCCTGCACATTACTAAAGTGGAAAATGTTAGGATTTTTCGGCAT	470				
QY	1117	CGCATGTTPAGAATCTCTAAAAATTTTAAACAATCTCTGTTTAAATGAATAAGTGTTGCTTTTAT	1176				
Db	469	CGCATGTTAGAATCTCTAAAAATTTTAAACAATCTCTGTTTAAATGAATAAGTGTTGCTTTTAT	410				
QY	1177	CAATATGAATTTCTGAAGGCCAAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAA	1236				
Db	409	CARTATGAATTTCTGAAGGCCAAATATCATACCATTAACTATGAAAGCTTTTAATTCCTAAA	350				
QY	1237	AATAGTTTTTAGAGATATTTCAAGCAAATGCTCTCTCTAATATCCAATAGCGAAGTGTTTATG	1296				
Db	349	AATAGTTTTTAGAGATATTTCAAGCAAATGCTCTCTCTAATATCCAATAGCGAAGTGTTTATG	290				
QY	1297	ACACAAATTCAGTACTGTGTTTAAAAATGAATCTTTATATTGACTGGGTGCCACAT	1356				
Db	289	ACACAAATTCAGTACTGTCT - GTTTAAAAATGAATCTTTTATTGACT - GGTGTTCCACAT	232				
QY	1357	ATTCCAGTAAATTTCTGTTATGAGAGCACTTGAATATAGCAAATTTGCCACACAGTTTAACTG	1416				
Db	231	ATTTTCAGTAAATTTCTGTTATGAGAGCACTTGAATATAGCAAATTTG - CCACACAGTTTAACTG	173				
QY	1417	GATAGATCACGTAGTGGTGTGATCATACCACTTGGTACTACACCCAGAACTCRAAATG	1476				
Db	172	GATAGACCACGTAGTGGTGTGATCATACCACTTGGTACTACACCCAGAACTCRAAATG	113				
QY	1477	TCTTTTCTCTTGATGAGATATGGGGTCTCTTTTGTACGTCTAGGCGCTAGGGTACCCAA	1536				
Db	112	TC - TTTCTCTGATGAGATAT - GGGTGTCTTTTGTACGTCTAGGCC - - TAGGTAAACA	58				
QY	1537	GTGAAATGAAATATATAPAGCAAAATGTGTTTGTATCCAGAGTCTTCTGTGCTGTGAAT	1594				
Db	57	GTGGAGTCATTTATATATAG - AAATGTGTTTGTATCCAGAGTCTTCTGTCTTGTGAAT	1				

172 GATAGACCAGCTACGCTGGTGATCATACCACTTGGTACTACCCAGAACTCAAATG 113

1477 TCTTTTCTCCTGATGAGATATGGGCTCTCTTTTGACGTCTAGGCTTAGGTAACCAA 1536

112 TC-TTTCCTCCTGATGAGATAT-GGGTCTCTTTTGTACGTCTAGGC---TAGGTAACCA 58

1537 GTGAAGTGAATATATATAGCAAAATGTCTTTGTATCCAGAGTCTTCCTGTCAATTGAAT 1594

57 GTGGAGTGATTATATATAGC-AAATGTCTTTGTATCCAGAGTCTTCCTGTCAATTGAAT 1

[illegible]

clone IMAGE:646753 3', mRNA sequence.

ACCESSION AA205598
VERSION AA205598.1 GI:1803606
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chapman, R., Cawthon, J., Cawthon, T., Cavallaro, A., Gibson, W.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PubMed 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine

TEL: 314 286 1800
Fax: 314 286 1810

Fax: 314 286 1810

Email: est@watson.wustl.edu
 This clone is available royalty-free through LINL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 345.

FEATURES
 source
 1..618
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:5217005"
 /db_xref="taxon:9606"
 /clone="IMAGE:646753"
 /dev_stage="Ntera-2/RA neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene neuroepithelium (#937231)"
 /notes="Vector: pBluescript SK-; Site:1. EcoRI; Site:2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2
 cells (Ntera-2/cl.dl) induced with Retinoic Acid for 24
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
 sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'"

ORIGIN
 Query Match 29.5%; Score 477.8; DB 9; Length 618;
 Best Local Similarity 93.6%; Pred. No. 8.3e-87;
 Matches 573; Conservative 0; Mismatches 30; Indels 9; Gaps 7;
 QY 1008 AACCAATATTGAATCTCTGATAAATATTTTCTGATGTTACTAGCTATGGAAATTAGAAC 1067
 DB 603 AACCCATATGGAACNCGGATAAATATTTTCGGANGTACTNGCTATGGAAATTTGGAAAC 544
 QY 1068 TGGCAACACCTGACATTAAGTCGGAATAGTTAGGATTTTCGGCATCGCATGTAGA 1127
 DB 543 TGGCACACCCCTGACATTAAGTCGGAATAGTTAGGATTTTCGGCATCGCATGTAGA 484
 QY 1128 ATCTCTAAATTTAAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAAT 1187
 DB 483 ATCTCTAAATTTAAACATTCCTGTTAAATGACTAAGGTTGCTTTTATCAATATGAAT 424
 QY 1188 CTGAAGGCCAATATCATACATTAATGAAGCTTTTAAATTCCTTAAATATAGTTTAA 1247
 DB 423 CTGAAGGCCAATATCATACATTAATGAAGCTTTTAAATTCCTTAAATATAGTTTAA 364
 QY 1248 AGATATTCAAGCAATGCTCTCTTAATATCCATAGCAAGTGTTTATGACACAAATCA 1307
 DB 363 AGATATTCAAGCAATGCTCTCTTAATATCCATAGCAAGTGTTTATGACACAAATCA 304
 QY 1308 CTAGTCTGGTTTAAATGAAATCTTTATATTGACTGGGTGCCACATATTTCCAGTAAT 1367
 DB 303 CTAGTCT-GTTTAAATGAAATCTTTATATTGACT-GGTGTTCCACATATTTCCAGTAAT 246
 QY 1368 TTCTGTTATGAGAGGACTTGAATAGCAATATGCCACACAGTTAACTGGATAGATCAG 1427
 DB 245 TTCTGTTATGAGAGGACTTGAATAGCAATATGCCACACAGTTAACTGGATAGATCAG 187
 QY 1428 TAGCTGGTATCATATACCACTTGGTACTACCCAGAAACTCAAAATTTGCTTTTCTCT 1487
 DB 186 TAGCTGGTATCATATACCACTTGGTACTACCCAGAAACTCAAAATTTGCT-TTCTCTCT 128
 QY 1488 GATGAGATATGGGTGCTCTTTTGTACGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGG 1547
 DB 127 GATGAGATAT-GGGTGTCTCTTTTGTACGCTAGGCTAGGCTAGGCTAGGCTAGG 72
 QY 1548 ATATAAGCAAAATGCTTTGTATCCAGAGCTTCTCTGCTCATTTGTAATAAAAAATTTAT 1607
 DB 71 ATATTAGC-ANATGCTTTGTATCCAGAGCTTCTCTGCTCATTTGTAATAAAAAATTTAT 13
 QY 1608 AAAAAATTTAAA 1619
 DB 12 AAAAAATTTAAA 1

RESULT 12

Z98443/c
 LOCUS Z98443 580 bp mRNA linear EST 22-SEP-1997
 DEFINITION HSZ98443 DKFZphtml Homo sapiens cDNA clone DKFZphtml_1e10 3',
 mRNA sequence.
 ACCESSION Z98443
 VERSION Z98443.1 GI:2326591
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 580)
 AUTHORS Korn,B., Wiemann,S., Ebert,L. and Poustka,A.
 TITLE Human ESTs (Korn,B. et al.)
 JOURNAL Unpublished (1997)
 COMMENT Contact: B. Korn
 Institution Molekulare Genomanalyse
 Deutsches Krebsforschungszentrum Heidelberg
 Im Neuenheimer Feld 506, D-69120 Heidelberg, FRG.
 FEATURES
 source
 1..580
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZphtml_1e10"
 /tissue_type="thymus"
 /dev_stage="adult"
 /clone_lib="DKFZphtml"
 ORIGIN
 Query Match 29.3%; Score 475; DB 14; Length 580;
 Best Local Similarity 95.7%; Pred. No. 3.1e-86;
 Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
 QY 1025 TGATAAATATTTTCTGATGTTTACTAGCTATGGAAATTAGAACTGGCAACCCCTGACAT 1084
 DB 580 TGATANATATTTCTGATGTTTACTAGCTATGGAAATTAGAACTGGCAACCCCTGACAT 521
 QY 1085 TACTAAGTGAATGTTAGGATTTTCGGCATCGCATGCTAGATCTCTAAATTTAAAC 1144
 DB 520 TACTAAGTGAATGTTAGGATTTTCGGCATCGCATGCTAGATCTCTAAATTTAAAC 461
 QY 1145 ATTCTCTTAAATGACTAAGGTTTGTCTTTTATCAATATGAATTCGAAGGCCAATATCAT 1204
 DB 460 ATTCTCTTAAATGACTAAGGTTTGTCTTTTATCAATATGAATTCGAAGGCCAATATCAT 401
 QY 1205 ACCATTAACTATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGAGATATTCAGCAATGC 1264
 DB 400 ACCATTAACTATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGAGATATTCAGCAATGC 341
 QY 1265 TCTCTTAATATCCATAGCAAGTGTTTATGACACAAATTCAGTCTCTGGTTTAAAAA 1324
 DB 340 TCTCTTAATATCCATAGCAAGTGTTTATGACACAAATTCAGTCTCT-GTTTAAAAA 282
 QY 1325 TGAATCTTTATATTGACTGGGTGCCACATATTCAGTAATTTCTGTTATGAGAGAC 1384
 DB 281 TGAATCTTTATATTGACT-GGTGTTCCACATATTTTCAGTAATTTCTGTTATGAGAGAC 223
 QY 1385 TTGAATAGCAAAATGCCACACAGTTAACTGGATAGATCACGTACGTGGTGATCATTAAC 1444
 DB 222 TTGAATAGCAAAATG-CCACACAGTTAACTGGATAGACCCAGTACGTGGTGATCATTAAC 164
 QY 1445 CACTTGGTACTACACCCAGAAACTCAAAATTTGCTTTTCTCTGATGAGATATGGGTGT 1504
 DB 163 CACTTGGTACTACACCCAGAAACTCAAAATTTGTC-TTCTCTCTGATGAGATAT-GGGTGT 106
 QY 1505 CCTTTTGTACGTCTAGGCTTAGGTTACCCAGTGAAGTGAATATATAAACAATATGTT 1564
 DB 105 CCTTTTGTACGTCTAGGCC---TAGGTAAACAGTGGAGTGAATATATATAGC-AAATGTGT 50
 QY 1565 TTGTATCCAGAGTCTTCTGTCTATTGTAATAAAAAATTTTAAAAA 1612
 DB 49 TTGTATCCAGAGTCTTCTGTCTATTGTAATAAAAAATTTTAAAAA 2

[illegible]

548 TTTACT-GCTTGAATACATTTGAGTTGAAGTTTGAAGAACTAATAGCAATTAATGA 606
181 TTTACTGCTTGAATACATTTGAGTTGAAGTTTGAAGAACTAATAGCAATTAATGA 240
607 ATAATGATGGAATTAATTCCTTGAAGAACTGATTAATTAATTAATTCCTTCTCT 666
241 ATAATGATGGAATTAATTCCTTGAAGAACTGATTAATTAATTAATTCCTTCTCT 300
667 TAGAAACAGTCAAAAGCCACTTTCAAAAGTTTCAAAATGAAGGAAGTACAAAGTTAGG 726
301 TAGAAACAGTCAAAAGCCACTTTCAAAAGTTTCAAAATGAAGGAAGTACAAAGTTAGG 360
727 CGATGATTAATTTCTGCTTGTGATACCACTGGCCAGGCTTTTATAGGACT 786
361 CGATGATTAATTTCTGCTTGTGATACCACTGGCCAGGCTTTTATAGGACT 418
787 CCCAAAGCATTTTGAAGATGGCAATATCAAAATGAAGTTGATGCTCTCTCAAAAGGCA 846
419 CCAAAAGCATTTTGAAGATGGCAATATCAAAATGAAGTTGATGCTCTCTCAAAAGGCA 478
847 TTTTAAATGTTACAA--TCTATTTGGAGCTCAGGTT-----ATGATATGTTATGA 900
479 TTTTAAATGTTTAAATCTAATTTGGAGCTCAGGTTTATGATATGTTTATGAACAA 538
901 ATAAGCTTCATTTTCTTATAGTACATCTCTTATTT--CCCTTTTGAACAAAGAA-T 957
539 TAAGCTTCATTTTCTTATAGTACATCTCTTATTTTCCCTTTTGAACAAAGAA-T 598
958 AACAATAGTTTAAATGTTGCAATTTAGTATTTAGTATTTAGTATTTAGTATTTAGTATTT 1017
599 ACCATTAGGTTTAAATGTTGCAATTTAGTATTTAGTATTTAGTATTTAGTATTTAGTATTT 658
1018 GAATCTCT--GATAATATTTCTGATGTTACTAGTATGGAATTTAGAACTGGCA 1072
659 TGAATCTCTGGTAAATTTTTCGGTGGGTACAGACTTGGGAATTAGAAATGGCA 716

RESULT 15
AW969584
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW969584 626 bp mRNA linear EST 01-JUN-2000
EST381661 WAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.
AW969584
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 626)
Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 267
Seq primer: Forward.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="WAGE resequences, MAGK"
/note="Vector: pBluescript-SKm"

FEATURES
source
ORIGIN

Query Match 28.5%; Score 461.2; DB 10; Length 626;
Best Local Similarity 94.5%; Pred. No. 1.9e-83; Indels 9; Gaps 7;
Matches 554; Conservative 0; Mismatches 23;
Qy 1028 TAAATATTTTCTCATGTTTACTAGCTATGGAATTTAGAACTGCGCATGTTAGAACTCTTAAAAATTTAAACATT 1147
Db 1 TAAATATTTTCTCATGTTTACTAGCTATGGAATTTAGAACTGCGCATGTTAGAACTCTTAAAAATTTAAACATT 120
Qy 1088 TAAGTGGAAATGTTAGGATTTTTCGGCATGCGCATGTTAGAACTCTTAAAAATTTAAACATT 1207
Db 61 TAAGTGGAAATGTTAGGATTTTTCGGCATGCGCATGTTAGAACTCTTAAAAATTTAAACATT 120
Qy 1148 CTTGTTTAAATGACTAAGGTTTGTCTTTTATCAATATGAATTTCTGAAGGCCAATATCATACC 1207
Db 121 CTTGTTTAAATGACTAAGGTTTGTCTTTTATCAATATGAATTTCTGAAGGCCAATATCATACC 180
Qy 1208 ATTAACATGAAAGCTTTTAAATTTCTTAAAAATTTTAAAGATATTTCAAGCAATGCTCT 1267
Db 181 ATTAACATGAAAGCTTTTAAATTTCTTAAAAATTTTAAAGATATTTCAAGCAATGCTCT 240
Qy 1268 CCTAATATCCATACGCAAGTGTCTTTTATGACACAAATTTCACTAGTCT-GTTTAAAAATGA 1327
Db 241 CCTAATATCCATACGCAAGTGTCTTTTATGACACAAATTTCACTAGTCT-GTTTAAAAATGA 299
Qy 1328 AATCTTATATGACTGGGTGTCCTCAATATTCAGTAATTTCTGTTATGAGAGGACTTG 1387
Db 300 AATCTTATATGACT-GGTGTTCCACATATTTCACTAGTCTTTCTGTTATGAGAGGACTTG 358
Qy 1388 AATAGCAAAATTTGCCACACAGTTTAACTGGATAGATCACCTAGCTGATCATACCCAC 1447
Db 359 AATAGCAAAATTTG-CCACACAGTTTAACTGGATAGACACCTAGCTGATCATACCCAC 417
Qy 1448 TTGGTACTACACCCAGAACTCAAAATTTGCTTTTCTCTGATGAGATATGGGGTCTCT 1507
Db 418 TTGGTACTACACCCAGAACTCAAAATTTGCT-TTTCTCTGATGAGATAT-GGGTGTCTCT 475
Qy 1508 TTTGCTAGTCTAGGGCTTAGGGTACCCAGTGAAGTGAATATATAGCAAAATTTGTTTG 1567
Db 476 TTTGCTAGTCTAGGGC---TAGGTAAACAGTGGAGTATATATAGC-AAATGTGTTTG 531
Qy 1568 TATCCAGAGTCTTCTCTGATTTGTAATAAAAATTTTATTTAAAAAT 1613
Db 532 TATCCAAAAGTCTCTGACATTTGTAATAAAAATTTTATTTAAAAAT 577

Search completed: April 26, 2004, 01:59:22
Job time : 2943.54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 80.3165 Seconds
(without alignments)
11186.557 Million cell updates/sec

Title: US-10-051-835-15

Perfect score: 1619
Sequence: 1 ttaactacatatctatgt.....atttatttaaaatttaaaa 1619

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2.6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2.6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2.6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2.6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2.6/prodata/2/ina/PCUS_COMB.seq:*
- 6: /cgn2.6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	59.8	3.7	7218	1	US-08-232-463-14
C 2	47.2	2.9	1664976	4	US-08-916-421B-1
C 3	46.4	2.9	339	4	US-09-821-976-8976
C 4	45.4	2.8	11049	4	US-10-204-708-22
C 5	44.6	2.8	1664976	4	US-08-916-421B-1
C 6	44.4	2.7	640681	4	US-09-790-988-1
C 7	44.4	2.7	8654	4	US-09-790-988-1
C 8	43.6	2.7	8654	1	US-08-920-812-6
C 9	43.6	2.7	8654	1	US-08-920-827-6
C 10	43.6	2.7	8654	1	US-08-921-177-6
C 11	43.6	2.7	8654	1	US-08-362-577C-6
C 12	43.6	2.7	8654	2	US-08-920-828-6
C 13	43.2	2.7	832	4	US-09-621-976-2813
C 14	43.2	2.7	1542	4	US-09-202-893B-3
C 15	42.8	2.6	832	4	US-09-621-976-2813
C 16	42.2	2.6	493	4	US-09-919-172-35
C 17	42.2	2.6	1500	4	US-08-601-198-36
C 18	40.8	2.5	6156	4	US-10-204-708-60
C 19	40.8	2.5	7198	4	US-09-453-702B-147
C 20	40.6	2.5	64467	4	US-09-803-671B-3
C 21	40	2.5	19124	2	US-08-487-826B-13
C 22	39.8	2.5	2861	1	US-08-299-953-1
C 23	39.8	2.5	2861	1	US-08-459-415-1
C 24	39.8	2.5	2861	1	US-09-866-687-1
C 25	39.8	2.5	2861	5	US-08-299-953-2
C 26	39.8	2.5	3881	1	US-08-299-953-2
C 27	39.8	2.5	3881	1	US-08-459-415-2

28	39.8	2.5	3881	4	US-09-866-687-2	Sequence 2, Appli
29	39.8	2.5	3881	5	PCT-US95-11231-2	Sequence 2, Appli
C 30	39.6	2.4	2614	4	US-09-004-056-1	Sequence 1, Appli
C 31	39.4	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appli
32	39.2	2.4	319608	4	US-09-539-333D-1	Sequence 1, Appli
33	39.2	2.4	319608	4	US-09-539-333D-1	Sequence 1, Appli
34	39	2.4	2169	4	US-09-434-408-3	Sequence 3, Appli
C 35	39	2.4	162450	4	US-09-345-882-1	Sequence 1, Appli
36	38.8	2.4	1189	1	US-08-307-591-2	Sequence 2, Appli
C 37	38.8	2.4	5332	4	US-09-801-861-3	Sequence 3, Appli
38	38.6	2.4	665	2	US-08-883-795A-36	Sequence 36, Appli
39	38.6	2.4	9347	4	US-10-204-708-35	Sequence 35, Appli
C 40	38.2	2.4	665	2	US-08-883-795A-36	Sequence 36, Appli
C 41	38.2	2.4	1149	4	US-09-601-198-151	Sequence 151, App
42	38.2	2.4	10607	1	US-08-078-090-3	Sequence 3, Appli
43	38	2.3	2332	4	US-08-956-171E-210	Sequence 210, App
44	38	2.3	3400	4	US-09-694-127-1	Sequence 1, Appli
45	38	2.3	168575	4	US-09-426-290-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/SOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)936-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.7%; Score 59.8; DB 1; Length 7218;

Best Local Similarity 2.6%; Pred. No. 5.5e-05;
Matches 10; Conservative 214; Mismatches 155; Indels 0; Gaps 0;

QY 21 ATATAAATTGCGCTAAATCCATAGAAAAAGCATATAAAGTATTAAAATGTGATATATA 80
 ||| ||| ||| ||
Db 1451 ATAGAAGATTGGTACRERRERRERRERRERRERRERRERRERRERRERRERRER 1392

QY 81 GCNNNNNNNNNNNNNNNNNNNNNATAGGGAAGTTCAAGTCACCTCOATTGAAGAAC 140
 : ::::: : :::
Db 1391 RRR 1332

QY 141 AVATCTCTGAGCATAAGCAGCAGCCTCAGGTCTTGCTGGGATGCAGTCGACAGGAGG 200
 : ::::: : :::
Db 1331 RRERRR 1272

QY 201 GGGAATTAAGAAAGAGACTATATAATTGAAAAGGATATAAGCATTAATAATATACA 260
 : ::::: : :::
Db 1271 RRR 1212

QY 261 TATATAGCTATATCTATGTATGTATCTAACACAGAAGTTCAAAGTCACITTC AATTAAAGAA 320
 : ::::: : :::
Db 1211 RRERRR 1152

QY 321 ACATTTTTGAGCATGGCACGACGCTCAGGTCTTATGCTGGGATGCAGTCGACAGGATG 380
 : ::::: : :::
Db 1151 RRR 1092

QY 381 GCGAAATTAGAAAAGAGAA 399
 : ::::: : :::
Db 1091 RRRRRRRRRRRRRRRRRRR 1073

RESULT 2
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)

```

/ NAME/KEY: misc_feature
/ LOCATION: (682442)..(682442)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (713652)..(713652)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (741684)..(741684)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (779455)..(779455)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (779676)..(779676)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (855539)..(855539)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (871619)..(871619)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1096846)..(1096846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1664854)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
/ US-08-916-421B-1

```

Qy	898	AAATAGCTTCATTATTCTTATAGCTACATCCTATTATTCCCTTTTAGAAACAAGAAAT	957
Db	317447	AACATCACCTTAATAATATTATCATAGACGTTTTTATATTTTTGCARACTCCCTT	317388
Qy	958	AACAATAAGTTTTAAATAGTTGCCATACCTATAGCATTTATCAGGCTTAATGAACCAATATT	1017
Db	317387	ATAAATTTTATTTTCATCAATTCCAATATTAGATTTTCTCTTATGGTTTTTGTCTATTCTT	317328
Qy	1018	GAATCTCTGATAATA	1033
Db	317327	CTTCTCCAAACAATA	317312

RESULT 3
 US-09-621-976-8976
 ; Sequence 8976, Application US/09621976
 ; Patent No. 5639063
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 8976
 ; LENGTH: 399
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-621-976-8976

RESULT 4
US-10-204-708-22
; Sequence 22, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PISPENBROCK, Christian

```

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 22
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-22

Query Match          2.8%; Score 45.4; DB 4; Length 11049;
Best Local Similarity 48.3%; Pred. No. 0.18;
Matches 156; Conservative 0; Mismatches 166; Indels 1; Gaps 1;

QY 796 ATTGTGAAGATGGCAATATCAATTAAGTGTATGCTCTCTCAATGAGCGATTATTTAATT 855
DB 254 ATTGTGAAGATGTTGTTTTTAAAAAATAAGTGAATGTTAAAGAAATTTTATGTAATT 313
QY 856 GTTCAACTATTGAGCGCTCAGTTATGATATGTTTATGAAAAATAAGCTTCATTATT 915
DB 314 AGTATAAATTTTGGAGGATAATTTGATAAT-TTATAGTCAAAAAATTTTAAATTT 372
QY 916 TCTTATAGTACATCTTATATTCCTTTTGAACAACAGATAACAATAAGTTTAAATAG 975
DB 373 TTGTAGTTTATTATTATTTTATGATATTTAGATATTTTAGAAAAATTTATGATATATA 432
QY 976 TTGCCATCTTAGCAATTTATCAGGTCTAATGAACCAATATGATCTCTGATAAATTT 1035
DB 433 TGAGAGTTATTATAGAAATGTTGTTTATATATGATAAATATATAAATTAATTTAAAG 492
QY 1036 TTCTGATGTTACTAGCTATGGAAATTTAGAACTGCGACACCCCTGACATTAAGTGG 1095
DB 493 TTTTATTAGTAGTTTATTTTAAAAAATAATTATAGTCGGCGCGGTGTTATATTTGT 552
QY 1096 AATGTTAGGATTTTCGGCATCG 1118
DB 553 AATTTTAGTATTTTGGGAGGTGCG 575

RESULT 5
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
```

```

; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
```



```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

Query Match      2.8%; Score 44.6; DB 4; Length 1664976;
Best Local Similarity 47.3%; Pred. No. 1.3;
Matches 134; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 461 ATAAACATACGAGTAAGATGTCAAATGAGTGTGACTCCCTTAAGTAGATTAAGTG 520
Db 484933 AAAAAATTAAACAGATAAAAAAAGCTAAATAATTAATATATATATATAGTTT 484992
Qy 521 TGCATCTCTTTGTTTCTTAAATAATGATTTTACTGCTTGAATTTACATTTGAGTTGAAGTT 580
Db 484993 ACACATTTTCAGCATCGTCGGTAATTTTGAACACATATTTTCCAAATATCACGTCATCTC 485052
Qy 581 TAGAACTAACATAGCATTAATTAATGAATATGATGAGAAATTTATTTATCTTTGAAACT 640
Db 485053 TATAAGCTAAACAACACATTTCTCAGCAACAACCCCTAATTTTAAAAATTTAAATATCATAT 485112
Qy 641 GATTGATAAATATATATCCCTCTCTTTAGAAACAGTCAAAAGCCACTTTCAACAAGTTTC 700
Db 485113 ATCCCTAAATTCATTAAGTCTTTTAAATAGATTCAAAAGGTACTTCTCTCAAAATTTA 485172
Qy 701 AAATAAAGGAAGTAGCAAGTTAGGCGATGGATTAATATTTTC 743
Db 485173 CAATTACAAGTCATATTTCTGGGATTTTATCTTTTATATTTTC 485215

RESULT 6
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRO
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Query Match      2.7%; Score 44.4; DB 4; Length 640681;
Best Local Similarity 46.2%; Pred. No. 1.1;
Matches 147; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 504 TAAAGTAGATTAAAGTGTGCATCTTTGTTTCTTAAATAATGATTTTACTGCTTGAATTT 563
Db 580418 TAAAAAATCTAGATTTGATTTTATTTTAAATAATGAATTTTACTTTTACATT 580477
Qy 564 ACATTTGAGTTGAGTTTAGAACTAACATAGCATTAATTAATGAATATGATGCAAGAAATT 623
Db 580478 GTATATGTTAAAAAATTAATTAATCAAAATAAAAAATTTATCTTTTATCAAGATAG 580537
```


US-08-920-827-6/c
; Sequence 6, Application US/08920827
; Patent No. 570375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical isolate SE-22
US-08-920-827-6

Query Match 2.7%; Score 43.6; DB 1; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 440 TTATAATTAATGATGATCTGAATAAACAATACAGGATAAGATGCTCAATGAGTGTGACT 499
Db 2309 TTAATAATGAGAAATTAAGAAACAATCTTTATTTTGAATATAGCAATGCTACATTGC 2250
QY 500 CCCCTTAAGTAGATTAAGTGTGATCTTTGTTTCCTTAAATATGATTTTACTGCTTGA 559
Db 2249 TAATTCAGATTAACCTTTAAAGATAATTTGTTTATAACACTTATTTGTTTACTTCT 2190
QY 560 AATTACATTTGAGTTGAGTTTAGAAACTAAGCAATAGCAATATATATGATATGCAATGGA 619
Db 2189 TAATTTTATTTAGTTATTAATTAACCTAAATAATATAGCAATTAATATATTTCTTTAAATTTA 2130
QY 620 AATTATATCTCTTTGAAACCTGATTTGATTAATATATTC 657
Db 2129 AATTAACTACTATATTATGACGACACTTTTCGTATTTTC 2092

RESULT 10
US-08-921-177-5/c
; Sequence 6, Application US/08921177

Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical isolate SE-22
US-08-921-177-6

Query Match 2.7%; Score 43.6; DB 1; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 440 TTATAATTAATGATGATCTGAATAAACAATACAGGATAAGATGCTCAATGAGTGTGACT 499
Db 2309 TTAATAATGAGAAATTAAGAAACAATCTTTATTTTGAATATAGCAATGCTACATTGC 2250
QY 500 CCCCTTAAGTAGATTAAGTGTGATCTTTGTTTCCTTAAATATGATTTTACTGCTTGA 559
Db 2249 TAATTCAGATTAACCTTTAAAGATAATTTGTTTATAACACTTATTTGTTTACTTCT 2190
QY 560 AATTACATTTGAGTTGAGTTTAGAAACTAAGCAATAGCAATATATATGATATGCAATGGA 619
Db 2189 TAATTTTATTTAGTTATTAATTAACCTAAATAATATAGCAATTAATATATTTCTTTAAATTTA 2130
QY 620 AATTATATCTCTTTGAAACCTGATTTGATTAATATATTC 657
Db 2129 AATTAACTACTATATTATGACGACACTTTTCGTATTTTC 2092

RESULT 11
US-08-362-577C-6/c
; Sequence 6, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE: Staphylococcus epidermidis
ORGANISM: Clinical Isolate SE-22
US-08-362-577C-6

Query Match 2.7%; Score 43.6; DB 1; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 440 TTATAATTAATGATGATCTGTAATAACATACACAGGATAAGATGTCAAATGAGTGTGACT 499
DB 2309 TTAATTAAGAAATAGAAACATCTTATTTTGTGATATAGCAATGCTACATGC 2250
QY 500 CCCTTAAAGTAGATAAAGTGTGCTTTGTTCTTAAATATGATTTTACTGCTGA 559
DB 2249 TAATTCAGATATAAATCTTTAAAGATAATTTGTTTATAAACACCTTATTTTACTTCT 2190
QY 560 AATTACATTTGAGTTGAAGTTTGAAGAACTAACATACGATTAATATGATATGATGCGAA 619
DB 2189 TAATTTTATTTAGTTATTAATTAATACTAAATAATATAGCATTAATAATATTTCTTTAAATTA 2130
QY 620 AATTATATCCCTTTGAAACCTGATTGATAATATATTC 657
DB 2129 AATTATATCTATATATGACGACACTTTCGTAATTTTC 2092

RESULT 12
US-08-920-828-6/c
Sequence 6, Application US/08920828
Patent No. 5853998
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8654 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE: Staphylococcus epidermidis
ORGANISM: Clinical Isolate SE-22
US-08-920-828-6

Query Match 2.7%; Score 43.6; DB 2; Length 8654;
Best Local Similarity 50.0%; Pred. No. 0.44;
Matches 109; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 440 TTATAATTAATGATGATCTGTAATAACATACACAGGATAAGATGTCAAATGAGTGTGACT 499
DB 2309 TTAATTAAGAAATAGAAACATCTTATTTTGTGATATAGCAATGCTACATGC 2250
QY 500 CCCTTAAAGTAGATAAAGTGTGCTTTGTTCTTAAATATGATTTTACTGCTGA 559
DB 2249 TAATTCAGATATAAATCTTTAAAGATAATTTGTTTATAAACACCTTATTTTACTTCT 2190
QY 560 AATTACATTTGAGTTGAAGTTTGAAGAACTAACATACGATTAATATGATATGATGCGAA 619
DB 2189 TAATTTTATTTAGTTATTAATTAATACTAAATAATATAGCATTAATAATATTTCTTTAAATTA 2130
QY 620 AATTATATCCCTTTGAAACCTGATTGATAATATATTC 657
DB 2129 AATTATATCTATATATGACGACACTTTCGTAATTTTC 2092

RESULT 13
US-09-621-976-2813
Sequence 2813, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET 054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21

Db 44 MCTYTWTRANKRRMMWKRSNSWMMAMGCMTRWAAAPMWRW 3

Search completed: April 26, 2004, 02:11:08
Job time : 88.3165 secs

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAATCTACATATCTTATGATATAAATATGCTAATCATAGAAAAAGGATATAAG 60
DB 1 TTAATCTACATATCTTATGATATAAATATGCTAATCATAGAAAAAGGATATAAG 60
QY 61 TATTAATATGATATATAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
DB 61 TATTAATATGATATATAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
QY 121 GTCACCTTCAATTCAGAAACATATCTGAGCATAGAGCAGCTCAGGTCTTATGGTGG 180
DB 121 GTCACCTTCAATTCAGAAACATATCTGAGCATAGAGCAGCTCAGGTCTTATGGTGG 180
QY 181 GATGAGTGGACAGGAGGGGAAATATGAAAGAGAACTATATAATTTGAAAAAGGGAT 240
DB 181 GATGAGTGGACAGGAGGGGAAATATGAAAGAGAACTATATAATTTGAAAAAGGGAT 240
QY 241 ATAAAGCATTAAATATATGATATATAGCTATATCTATGATGATCTTAACAGAGATTC 300
DB 241 ATAAAGCATTAAATATATGATATATAGCTATATCTATGATGATCTTAACAGAGATTC 300
QY 301 AAGTCACTTCAATTTAAAGAACTATTTTGGACATGGGACAGCTCAGGTCTTATGCTGG 360
DB 301 AAGTCACTTCAATTTAAAGAACTATTTTGGACATGGGACAGCTCAGGTCTTATGCTGG 360
QY 361 GATGAGTGGACAGGAGGGGAAATATGAAAGAGAACTGTAATTTGAAATGACGTG 420
DB 361 GATGAGTGGACAGGAGGGGAAATATGAAAGAGAACTGTAATTTGAAATGACGTG 420
QY 421 GGCTGACACCTTAAAGAACTTAAATTAATGATGATCTGAATAAATACACAGGATAAG 480
DB 421 GGCTGACACCTTAAAGAACTTAAATTAATGATGATCTGAATAAATACACAGGATAAG 480
QY 481 ATGTCAAATGAGTGTGACTCCCTTAAAGTAGATTAAGTGTGCAATCTTTGTTCTCAA 540
DB 481 ATGTCAAATGAGTGTGACTCCCTTAAAGTAGATTAAGTGTGCAATCTTTGTTCTCAA 540
QY 541 ATATGATTTTACTGCTTGAATTTACATTTGAGTTGAAGTTTGAATTAACATAGCATTA 600
DB 541 ATATGATTTTACTGCTTGAATTTACATTTGAGTTGAAGTTTGAATTAACATAGCATTA 600
QY 601 ATATGAATATGATGAAATTTATATCTTTGAAACCTGATGATGATATATATTCCTCC 660
DB 601 ATATGAATATGATGAAATTTATATCTTTGAAACCTGATGATGATATATATTCCTCC 660
QY 661 CTCCTTTAGAAACAGTCAAAGCCACTTCAAACAGTTTCAAATTAAGGAGGTAGCAA 720
DB 661 CTCCTTTAGAAACAGTCAAAGCCACTTCAAACAGTTTCAAATTAAGGAGGTAGCAA 720
QY 721 GTTAGGGATGATATATTTCTTGCTGTTGTTATACCCATGGCCAGGGCCTTTATA 780
DB 721 GTTAGGGATGATATATTTCTTGCTGTTGTTATACCCATGGCCAGGGCCTTTATA 780
QY 781 AGGACTCCCAAGCACTTTTGAAGATGCAATATCAATTAAGTGTATGCTCTCAAT 840
DB 781 AGGACTCCCAAGCACTTTTGAAGATGCAATATCAATTAAGTGTATGCTCTCAAT 840
QY 841 GAGCACTTTTAAATGTTAACTATTTGAGAGCTCAGGTATGATATGTTTATGAAA 900
DB 841 GAGCACTTTTAAATGTTAACTATTTGAGAGCTCAGGTATGATATGTTTATGAAA 900
QY 901 ATAGCTTCATATTTCTTATAGTACATCTTATATTTCCCTTTTGAACAGAGATAAC 960
DB 901 ATAGCTTCATATTTCTTATAGTACATCTTATATTTCCCTTTTGAACAGAGATAAC 960
QY 961 AATAAGTTTAAATAGTTGCCATCTTAGCATTTATATGAGTCTAATGAAACCAATATGAA 1020
DB 961 AATAAGTTTAAATAGTTGCCATCTTAGCATTTATATGAGTCTAATGAAACCAATATGAA 1020
QY 1021 TCTCTGATTAATATTTCTGATGTTACTAGCTATGGGAATTAAGTGTGCAACCCCTG 1080

DB 1021 TCTCTGATTAATATTTTCTGATGTTACTAGTATGGAAATTAAGAACTGSCAACCCCTG 1080
QY 1081 ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTT 1140
DB 1081 ACATTACTAAGTGGAAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTT 1140
QY 1141 AAACATTCCTGTTAAATGACTTAAGGTTTGTCTTTATCAATATCAATTTCTGAAGGCCAATA 1200
DB 1141 AAACATTCCTGTTAAATGACTTAAGGTTTGTCTTTATCAATATGATTTCTGAAGGCCAATA 1200
QY 1201 TCATACCATTTAACTATGAAAGCTTTTAAATTCCTTAAATAAGTTTATAGATATTCAGCA 1260
DB 1201 TCATACCATTTAACTATGAAAGCTTTTAAATTCCTTAAATAAGTTTATAGATATTCAGCA 1260
QY 1261 ATGCTCTCTTAAATATCCATAGCAAGTGTCTTTATGACACAAATTCACATAGTCTGTTTA 1320
DB 1261 ATGCTCTCTTAAATATCCATAGCAAGTGTCTTTATGACACAAATTCACATAGTCTGTTTA 1320
QY 1321 AAAATGAAATCTTTTATATGACTGGGTGTCCTCATATTCACGTAATTTCTGTTATGAGA 1380
DB 1321 AAAATGAAATCTTTTATATGACTGGGTGTCCTCATATTCACGTAATTTCTGTTATGAGA 1380
QY 1381 GGACTTGAATAGCAAAATTCGCCACAGTTAACTGGATAGATCACGTCGTGTATCA 1440
DB 1381 GGACTTGAATAGCAAAATTCGCCACAGTTAACTGGATAGATCACGTCGTGTATCA 1440
QY 1441 TAACCACTTGGTACTACACCCAGAACTCAAAATTTGTCTTTCTCTGATGAGATATGG 1500
DB 1441 TAACCACTTGGTACTACACCCAGAACTCAAAATTTGTCTTTCTCTGATGAGATATGG 1500
QY 1501 GTGTCTTTTGTAGCTTAGGGCTAGGGTACCCAGTGAAGTGAATATATAGCAAAAT 1560
DB 1501 GTGTCTTTTGTAGCTTAGGGCTAGGGTACCCAGTGAAGTGAATATATAGCAAAAT 1560
QY 1561 GTGTCTTGTATCCAGAGTCTTCTCTGATGATTAATAAAAAATTTATTTAAAAATTTAAA 1619
DB 1561 GTGTCTTGTATCCAGAGTCTTCTCTGATGATTAATAAAAAATTTATTTAAAAATTTAAA 1619

RESULT 2
ABX74725/c

ID ABX74725 standard; cDNA; 618 BP.

XX AC ABX74725;

XX XX 21-MAR-2003 (first entry)

XX Human cDNA sequence #135 down-regulated in CC-RCC patients.

XX Human; microarray; solid surface; immobilised probe; CC-RCC;

KW differential expression profile; aggressive CC-RCC tumour type;

KW non-aggressive CC-RCC tumour type; clear cell renal carcinoma;

KW gene expression profiling; tumour tissue; gene; ss.

XX Homo sapiens.

OS WO200279411-A2.

XX PN 10-OCT-2002.

XX PD 29-MAR-2002; 2002WO-US009576.

XX PF 29-MAR-2001; 2001US-0279411P.

XX PR (VAND-) VAN ANDEL INST.

XX PA Haab B, Rhodes D, Teh BT, Takashi M;

XX PI WPI; 2003-040679/03.

XX DR New microarray, comprising a matrix of cDNA probe from a set of probes

XX PT immobilized to a solid surface in predetermined order, useful in the

PT prognosis of patients with clear cell renal carcinoma.

XX Example 3; SEQ ID NO 295; 179pp; English.

PS The present invention relates to a microarray comprising a matrix of at

XX least one cDNA probe from a set of probes immobilised to a solid surface

CC in a predetermined order, where a row of pixels corresponds to replicates

CC of one distinct probe from the set. The probes are complementary to

CC nucleic acid sequences that are expressed differentially in aggressive as

CC compared to non-aggressive types of clear cell renal carcinoma (CC-RCC)

CC and that hybridise to the probes under high stringency conditions. The

CC microarray is useful for the prognosis of patients with CC-RCC, wherein

CC aggressive and non-aggressive CC-RCC tumour types are characterised by

CC differential expression profiles of genes that hybridise with one or more

CC probes immobilised on the microarray. The arrays are useful for gene

CC expression profiling of tumour and normal tissues. The present sequence

XX represents a human cDNA sequence down-regulated in CC-RCC patients

SQ Sequence 618 BP; 203 A; 106 C; 106 G; 197 T; 0 U; 6 Other;

Query Match 29.5%; Score 477.8; DB 7; Length 618;

Best Local Similarity 93.6%; Pred. No. 1.2e-95;

Matches 573; Conservative 0; Mismatches 30; Indels 9; Gaps 7;

QY 1008 AACCATATGATCTCTGATTAATATTTTCGATGTTACTAGCTATCGGAAATTTAGAAC 1067

Db |||||||

QY 603 AACCCATATGGACCCGCGATTAATTTTCGGANGTACTGCTATGGAAATTTGGAAC 544

Db |||||||

QY 1068 TGGCACAACCCCTGACATTAAGTGGAAATTTAGGATTTTCGGCATCGCATGTTAGA 1127

Db |||||||

QY 543 TGGCACAACCCCTGACATTAAGTGGAAATTTAGGATTTTCGGCATCGCATGTTAGA 484

Db |||||||

QY 1128 ATCTCTAAATTTAAACATCTCTGTTAAATGACTAAGGTTTCTTTTCAATATGAAT 1187

Db |||||||

QY 483 ATCTCTAAATTTAAACATCTCTGTTAAATGACTAAGGTTTCTTTTCAATATGAAT 424

Db |||||||

QY 1188 CTGAAGGCCAATATCATACCATTAACCTATGAAGCTTTTAAATCTCTAAATATGTTTAG 1247

Db |||||||

QY 423 CTGAAGGCCAATATCATACCATTAACCTATGAAGCTTTTAAATCTCTAAATATGTTTAG 364

Db |||||||

QY 1248 AGATATTCAGCATCTCTCTATATCATACGCAAGTGTTTATGACACAAATTTCA 1307

Db |||||||

QY 363 AGATATTCAGCAATCTCTCTATATCATACGCAAGTGTTTATGACACAAATTTCA 304

Db |||||||

QY 1308 CTAGTCTGTTTAAATGAATCTTTATATGACTGGGTGTCACACATATTCAGTAAAT 1367

Db |||||||

QY 303 CTAGTCT-GTTTAAATGAATCTTTATATGACT-GGTGTTCCACATATTTCAAT 246

Db |||||||

QY 1368 TTCTGTTATGAGGACCTTGAATAGCAATTTGCCACACAGTTAACTGGATAGATCAG 1427

Db |||||||

QY 245 TTCTGTTATGAGGACCTTGAATAGCAATTTGCCACACAGTTAACTGGATAGATCAG 187

Db |||||||

QY 1428 TAGCTGTTATGATCAATACCACTTGGTACTACACCGAAGACTCAAAATCTCTTTCTCT 1487

Db |||||||

QY 186 TAGCTGTTATGATCAATACCACTTGGTACTACACCGAAGACTCAAAATCTCTTTCTCT 128

Db |||||||

QY 1488 GATGAGATATGGGCTGCTCTTTTGTACGTTAGGCTAGGGCTAGGGTACCCAGTGAAT 1547

Db |||||||

QY 127 GATGAGATAT-GGGTGTCTCTTTTGTACGTTAGGCTAGGGCTAGGGTGAAT 72

Db |||||||

QY 1548 ATATAGCAAAATGCTTTGTTATCCAGACTCTCTGCTCATTTGTAATAAAATTTATTT 1607

Db |||||||

QY 71 ATATAGCAAAATGCTTTGTTATCCAGACTCTCTGCTCATTTGTAATAAAATTTATTT 13

Db |||||||

QY 1608 AAAAAATTTAAAA 1619

Db |||||||

QY 12 AAAAAATTTAAAA 1

Db |||||||

RESULT 3

ABL68031/c

ID ABL68031 standard; DNA; 580 BP.

XX

AC ABL68031;

XX 15-MAY-2002 (first entry)

DT Ovary cancer related gene sequence SEQ ID NO:6368.

DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

XX gene; ds

OS Homo sapiens.

XX WO200194629-A2.

EN 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US010838.

XX 05-JUN-2000; 2000US-0209473P.

PR 05-JUN-2000; 2000US-0209531P.

PR 18-SEP-2000; 2000US-0233133P.

PR 18-SEP-2000; 2000US-0233617P.

PR 20-SEP-2000; 2000US-0234009P.

PR 20-SEP-2000; 2000US-0234034P.

PR 20-SEP-2000; 2000US-0234052P.

PR 22-SEP-2000; 2000US-0234509P.

PR 22-SEP-2000; 2000US-0234567P.

PR 25-SEP-2000; 2000US-0234923P.

PR 25-SEP-2000; 2000US-0235077P.

PR 25-SEP-2000; 2000US-0235082P.

PR 25-SEP-2000; 2000US-0235134P.

PR 25-SEP-2000; 2000US-0235280P.

PR 26-SEP-2000; 2000US-0235637P.

PR 26-SEP-2000; 2000US-0235838P.

PR 27-SEP-2000; 2000US-0235711P.

PR 27-SEP-2000; 2000US-0235720P.

PR 27-SEP-2000; 2000US-0235840P.

PR 28-SEP-2000; 2000US-0235863P.

PR 28-SEP-2000; 2000US-0236028P.

PR 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.

PR 28-SEP-2000; 2000US-0236034P.

PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.

PR 29-SEP-2000; 2000US-0236842P.

PR 29-SEP-2000; 2000US-0236891P.

PR 02-OCT-2000; 2000US-0237172P.

PR 02-OCT-2000; 2000US-0237173P.

PR 02-OCT-2000; 2000US-0237278P.

PR 02-OCT-2000; 2000US-0237294P.

PR 02-OCT-2000; 2000US-0237295P.

PR 02-OCT-2000; 2000US-0237316P.

PR 03-OCT-2000; 2000US-0237425P.

PR 03-OCT-2000; 2000US-0237598P.

PR 03-OCT-2000; 2000US-0237604P.

PR 03-OCT-2000; 2000US-0237606P.

PR 03-OCT-2000; 2000US-0237608P.

PR 01-NOV-2000; 2000US-0244867P.

PR 01-NOV-2000; 2000US-0245084P.

XX (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

PI Soppet DR, Weaver Z;

XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a chemical

PT agent to be tested for anti-neoplastic activity, and determining a change

PT in expression of a gene of a signature gene set.

XX Claim 1; SEQ ID NO 6368; 44pp; English.

PS

CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (II) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'
 CC tumour

XX SQ Sequence 580 BP; 198 A; 96 C; 96 G; 187 T; 0 U; 1 Other;
 Query Match 29.3%; Score 475; DB 6; Length 580;
 Best Local Similarity 95.7%; Pred. No. 5e-95;
 Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
 QY 1025 TGATAATATTTCTGATGTTACTAGTATGGGAATAGAACTGGCAACCCCTGACAT 1084
 Db 580 TGATAATATTTCTGATGTTACTAGTATGGGAATAGAACTGGCAACCCCTGACAT 521
 QY 1085 TACTAAGTGAATGTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTAAAC 1144
 Db 520 TACTAAGTGAATGTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTAAAC 461
 QY 1145 ATTCTGTTAAATGACTAGGTTGTTTTCATATGAAATCTGAAGGCCAATATCAT 1204
 Db 460 ATTCTGTTAAATGACTAGGTTGTTTTCATATGAAATCTGAAGGCCAATATCAT 401
 QY 1205 ACCATTAACTATCAAGAGCTTTTAAATCCTTAAATAATGTTTAGAGATTTCAAGCAATGC 1264
 Db 400 ACCATTAACTATCAAGAGCTTTTAAATCCTTAAATAATGTTTAGAGATTTCAAGCAATGC 341
 QY 1265 TCTCTAATATCCATACGCAAGTGTTTATGACACAAATTCACATGCTGTTTAAAAA 1324
 Db 340 TCTCTAATATCCATACGCAAGTGTTTATGACACAAATTCACATGCTGTTTAAAAA 282
 QY 1325 TGAATCTTTATATTGACTGGGTGTCACATATCCAGTAATTTCTGTTATGAGAGAC 1384
 Db 281 TGAATCTTTATATTGACT--GGTGTCCACATATTCAGTAATTTCTGTTATGAGAGAC 223
 QY 1385 TTGAATAGCAAAATGTCACACAGTTAATCTGATAGATCAAGTACGTTGGTATCATTAAC 1444
 Db 222 TTGAATAGCAAAATG--CCACAGATTAACTGGATAGACCAGTACGTTGGTATCATTAAC 164
 QY 1445 CACTTGTGTTACTACACCCAGAACTCAAAATTTGTTCTTCTGATGATATGGGTGT 1504
 Db 163 CACTTGTGTTACTACACCCAGAACTCAAAATTTGTTCTTCTGATGATAT--GGGTGT 106
 QY 1505 CTTTGTGTTAGCTAGGGCTAGGGTACCAAGTGAAGTGAATATATAGCAAAATGTT 1564
 Db 105 CTTTGTGTTAGCTAGGGCTAGGGTACCAAGTGAAGTGAATATATAGCAAAATGTT 50
 QY 1565 TTGTATCCAGAGCTTCTGCTCATCTGATATATAAAATTTTATTTAAAAA 1612
 Db 49 TTGTATCCAGAGCTTCTGCTCATCTGATATATAAAATTTTATTTAAAAA 2

RESULT 5

ABL67046/c

ID ABL67046 standard; DNA; 580 BP.

XX AC ABL67046;

XX AC

XX 15-MAY-2002 (first entry)

XX DE

Thyroid cancer related gene sequence SEQ ID NO:5383.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
 KW gene; ds.
 XX Homo sapiens.
 OS WO200194629-A2.
 PN 13-DEC-2001.
 PD 30-MAY-2001; 2001WO-US010838.
 PF 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX (AVAL-) AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a chemical
 agent to be tested for anti-neoplastic activity, and determining a change
 in expression of a gene of a signature gene set.

Claim 1; SEQ ID NO 5383; 4pp; English.

The present invention describes a method (M1) for screening for an anti-
 neoplastic agent. The method involves exposing cells to a chemical agent
 to be tested for anti-neoplastic activity, determining a change in

expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour

XX Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 5e-95;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;

QY 1025 TGATAAATATTTCTGATGTTACTAGCTATGCGCATCGCATGTTAGAACTCTTAAATTTAAAC 1084
Db 580 TGATANATATTTCTGATGTTACTAGCTATGCGCAATTTAGAACTCTTAAATTTAAAC 521

QY 1085 TACTAGTGAATGTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTAAAC 1144
Db 520 TACTAGTGAATGTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTAAAC 461

QY 1145 ATTCTGTTAAATGATGATAGTTGTTTATCAATGATGATGTTGATGATGATGATGATGATGAT 1204
Db 460 ATTCTGTTAAATGATGATAGTTGTTTATCAATGATGATGATGATGATGATGATGATGAT 401

QY 1205 ACCATTAACTATCAAGCTTTTAAATCTTAAATTTAGATATTTCAAGCAATGC 1264
Db 400 ACCATTAACTATCAAGCTTTTAAATCTTAAATTTAGATATTTCAAGCAATGC 341

QY 1265 TCTCCTAATATCCATAGCGAAGTGTGTTTATGACACAAATTTCTAGTCTGGTTTAAAAA 1324
Db 340 TCTCCTAATATCCATAGCGAAGTGTGTTTATGACACAAATTTCTAGTCTGGTTTAAAAA 282

QY 1325 TGAATCTTTATATGCTAGCTGGGTGTCACATATTCAGTAATTTCTGTTATGAGAGGAC 1384
Db 281 TGAATCTTTATATGCTAGCTGGGTGTCACATATTCAGTAATTTCTGTTATGAGAGGAC 223

QY 1385 TTGAATAGCAAAATGCCCCACACAGTTAACTGGATAGATCACATGCTGGTGTATCAATAC 1444
Db 222 TTGAATAGCAAAATGCCCCACACAGTTAACTGGATAGATCACATGCTGGTGTATCAATAC 164

QY 1445 CACTTGTACTACACCCAGAACTCAAAATGTTCTTTCTCTGATGATGATGGGTGT 1504
Db 163 CACTTGTACTACACCCAGAACTCAAAATGTTCTTTCTCTGATGATGATGGGTGT 106

QY 1505 CCTTTTGTAGCTCTAGGCGCTTAGGCTACCCAAAGTGAAGTGAATATATATAGCAAAATGTGT 1564
Db 105 CCTTTTGTAGCTCTAGGCGCTTAGGCTACCCAAAGTGAAGTGAATATATATAGCAAAATGTGT 50

QY 1565 TTGATCCAGAGTCTTCTGTCATGTAATTAATAAATTTTAAATAA 1612
Db 49 TTGATCCAGAGTCTTCTGTCATGTAATTAATAAATTTTAAATAA 2

RESULT 6
ABL64323/C
ID ABL64323 standard; DNA; 580 BP.
AC ABL64323;
XX
XX
XX 15-MAY-2002 (first entry)
XX
XX Stomach cancer related gene sequence SEQ ID NO:2660.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
OS Homo sapiens.
XX W0200194629-A2.
XX 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US010838.
XX 05-JUN-2000; 2000US-0209473P.
XX 05-JUN-2000; 2000US-0209531P.
XX 18-SEP-2000; 2000US-0231333P.
XX 18-SEP-2000; 2000US-0233617P.
XX 20-SEP-2000; 2000US-0234009P.
XX 20-SEP-2000; 2000US-0234034P.
XX 20-SEP-2000; 2000US-0234052P.
XX 22-SEP-2000; 2000US-0234509P.
XX 22-SEP-2000; 2000US-0234567P.
XX 25-SEP-2000; 2000US-0234923P.
XX 25-SEP-2000; 2000US-0234924P.
XX 25-SEP-2000; 2000US-0235077P.
XX 25-SEP-2000; 2000US-0235082P.
XX 25-SEP-2000; 2000US-0235134P.
XX 25-SEP-2000; 2000US-0235280P.
XX 26-SEP-2000; 2000US-0235637P.
XX 26-SEP-2000; 2000US-0235638P.
XX 27-SEP-2000; 2000US-0235711P.
XX 27-SEP-2000; 2000US-0235720P.
XX 27-SEP-2000; 2000US-0235840P.
XX 27-SEP-2000; 2000US-0235863P.
XX 28-SEP-2000; 2000US-0236028P.
XX 28-SEP-2000; 2000US-0236032P.
XX 28-SEP-2000; 2000US-0236033P.
XX 28-SEP-2000; 2000US-0236034P.
XX 28-SEP-2000; 2000US-0236109P.
XX 28-SEP-2000; 2000US-0236111P.
XX 29-SEP-2000; 2000US-0236842P.
XX 29-SEP-2000; 2000US-0236891P.
XX 02-OCT-2000; 2000US-0237172P.
XX 02-OCT-2000; 2000US-0237173P.
XX 02-OCT-2000; 2000US-0237278P.
XX 02-OCT-2000; 2000US-0237294P.
XX 02-OCT-2000; 2000US-0237295P.
XX 02-OCT-2000; 2000US-0237316P.
XX 03-OCT-2000; 2000US-0237425P.
XX 03-OCT-2000; 2000US-0237598P.
XX 03-OCT-2000; 2000US-0237604P.
XX 03-OCT-2000; 2000US-0237606P.
XX 03-OCT-2000; 2000US-0237608P.
XX 01-NOV-2000; 2000US-0244867P.
XX 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PB, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, and determining a change
XX in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 2660; 44pp; English.
XX The present invention describes a method (M1) for screening for an anti-
XX neoplastic agent. The method involves exposing cells to a chemical agent
XX to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664

CC to ABL70110, or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
CC tumour
XX
SQ Sequence 580 BP; 198 A; 98 C; 96 G; 187 T; 0 U; 1 Other;

Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 5e-95;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;

QY 1025 TGATAAATATTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTGACAT 1084
DB 580 TGATANATATTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTGACAT 521
QY 1085 TACTAGTGGAATGTTAGGATTTTCGGCATCGCATGTAGAACTCTTAAATTTAAAC 1144
DB 520 TACTAGTGGAATGTTAGGATTTTCGGCATCGCATGTAGAACTCTTAAATTTAAAC 461
QY 1145 ATTCCTGTTAAATGACTAAAGTTTGCTTTTATCAATATGAATTTCTGAAGCCCAATATCAT 1204
DB 460 ATTCCTGTTAAATGACTAAAGTTTGCTTTTATCAATATGAATTTCTGAAGCCCAATATCAT 401
QY 1205 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATATAGTTTATAGATATTCAGCAATGC 1264
DB 400 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATATAGTTTATAGATATTCAGCAATGC 341
QY 1265 TCTCCTAAATCATCATCGCAAGTGTTTATGACACAAATTCCTAGTCTGGTTTAAAAA 1324
DB 340 TCTCCTAAATCATCATCGCAAGTGTTTATGACACAAATTCCTAGTCTGGTTTAAAAA 282
QY 1325 TGAATCTTTATATGACTGGGTGCCACATATTCAGTAATTTCTGTTATGAGAGGAC 1384
DB 281 TGAATCTTTATATGACT--GGTGTCCACATATTTTCAGTAATTTCTGTTATGAGAGGAC 223
QY 1385 TTGAAATAGCAATGCGCCACACAGTTTAACTGATGATGATGATGATGATGATGATGATGAT 1444
DB 222 TTGAAATAGCAATG--CCACAGTTTAACTGATGATGATGATGATGATGATGATGATGAT 164
QY 1445 CACTTGTGTTACTACACCCAGAACTCAAAATTTGCTTTTCTCCTGATGATGATGATGATGAT 1504
DB 163 CACTTGTGTTACTACACCCAGAACTCAAAATTTGCTTTTCTCCTGATGATGATGATGATGAT 106
QY 1505 CTTTGTGATGCTAGGGCTAGGGTACCCAGTGAATGATGATGATGATGATGATGATGATGAT 1564
DB 105 CTTTGTGATGCTAGGGCTAGGGTACCCAGTGAATGATGATGATGATGATGATGATGATGAT 50
QY 1565 TTGTATCCAGAGTCTTCTGTCATTTGTAATAAAAAATTTATTTAAAAA 1612
DB 49 TTGTATCCAGAGTCTTCTGTCATTTGTAATAAAAAATTTATTTAAAAA 2

RESULT 7
ABL33246
ID ABL33246 standard; DNA; 6077 BP.
XX ABL33246;
AC
XX
AC
XX
DT 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1219.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; anti-anaemic; cytostatic; neutropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX
OS Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP007537.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX Claim 1; SEQ ID NO 1219; 32bp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6077 BP; 1524 A; 185 C; 1391 G; 2977 T; 0 U; 0 Other;

Query Match 3.2%; Score 52.6; DB 6; Length 6077;
Best Local Similarity 44.9%; Pred. No. 0.098;
Matches 199; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 809 GCAATATCAATAGTGTATGTCCTCTCAATGAGGCAATTTTAAATTTGTACAACTATT 868
DB 39 GCAATATTTAGAACTGATGATGTTAGTTTAAATTTTATTTATTTTGT 98
QY 869 TGGACGCTCAGGTTATGATATGTTTATGAAAAAAGCTTCATTATTTCTTATAGCTACA 928
DB 99 GAAAAAGTTTATTTTGTATATAATTTAAATTTTAAATTTTGTGGAGTTAAA 158
QY 929 TCCTATTATTCCTTTTGAACCAAGATAACAATAGTTTAAATAGTTGCCATCTTAG 988
DB 159 AATTATTTTAAATAGTTTATGTTTAAATGATATAAATTTTGTTCGTAAATTTT 218
QY 989 CATTTATCAGGCTTAATGAAACCAATATTTGAATCTTGATATAAATTTTCTGATGTTACT 1048
DB 219 ATTTATATTTTATTTTGAATAGTTTATTTAAATTTTATTTAGGTTTAAATTTGAGATGTT 278
QY 1049 ACCTATGGGAATTAGAACTGCGCAACCCCTGACATTAAGTGAATGTTAGGATTT 1108
DB 279 AAGTATAAATTTATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 338
QY 1109 TTCCGCATCGCATGTTAGAAATCTCTAAATTTTAAACATTTCTGTTTAAATGACTAGGTTT 1168
DB 339 TTTGATTTTGTGTTTGTGTTTATTTATTTATTTAGATTTTAAATTTTATTTTAAAAAT 398
QY 1169 GCTTTTATCAATATGAAATTCGAGGCCCAATATCATACCATTAAGTGAAGCTTTTAA 1228
DB 399 TTATATTTTGAATAAATTTATATATTTATTTATTTTGGAGATTTATGGATGATATAA 458
QY 1229 TTCTTAAAAATAGTTTATAGAT 1251

Haematopoietic cell proliferation disorder related DNA sequence #386.

PA (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-090045/12.
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX Claim 1; SEQ ID NO 281; 32pp; English.
PS The invention relates to a nucleic acid, which comprises a segment of the
XX chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX SQ Sequence 9504 BP; 2528 A; 138 C; 1853 G; 4985 T; 0 U; 0 Other;

Query Match 3.1%; Score 50.6; DB 6; Length 9504;
Best Local Similarity 53.9%; Pred. No. 0.3;
Matches 104; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 846 ATTTTAAATGTTTCAATCTATTGTCAGGCTTCAGGTTATGATATGTTTATGAAATAAG 905
DB 7045 ATTTTAAATGTTTCAATCTATTGTTTATCACTAAATTAATTTCCATATAATTAATC 6986

QY 906 CTCATATTTCTTATAGTACATCTCTATTATTTCCCTTTTAGAAACAGAAATACAAATA 965
DB 6985 ATACATTTATTAATTTCTTAAACTTTCATATTTTATTAATTAATTAATTAATTAATTA 6926

QY 966 GTTTTAAATGTTGCTACTTATGATTTATCAGGCTCTAATGAACCAATATTGCAATCTCT 1025
DB 6925 TTTCAATATCTCTTATTTTAAATAATCTATATTTAAATAATTAATTAATTAATTAATTA 6866

QY 1026 GATAAATATTTTC 1038
DB 6865 AAAAAATATTTTC 6853

RESULT 11
ABQ67093/c
ID ABQ67093 standard; DNA; 83391 BP.
XX AC ABQ67093;
XX 28-AUG-2002 (first entry)
XX DE Human angiogenesis associated polynucleotide SEQ ID NO 123.
XX KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antidiabetic;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiarteriosclerotic; ds.
XX OS Homo sapiens.
XX WO200246454-A2.
XX 13-JUN-2002.
XX 06-DEC-2001; 2001WO-EP014320.
XX 06-DEC-2000; 2000DE-01061338.
XX (EPIG-) EPIGENOMICS AG.
XX Schacht O;
XX WPI; 2002-500450/53.
XX New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX Claim 1; SEQ ID NO 123; 41pp + Sequence Listing; German.
XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 83391 BP; 25113 A; 665 C; 16761 G; 40835 T; 0 U; 17 Other;

Query Match 3.1%; Score 50.4; DB 6; Length 83391;
Best Local Similarity 43.8%; Pred. No. 0.53;
Matches 266; Conservative 0; Mismatches 341; Indels 1; Gaps 1;

QY 428 CCCTTAAAGAACTTATTAATTAATGATGATCTGAATAAACATACCAGGATTAAGATGCAA 487
DB 19871 CTCITTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19812

QY 488 ATGAGTGTGACTCCCTTAAAGTAGATTAAAGTGTGCAATCTTTTGTTCCTTAATAATGAT 547
DB 19811 CAACATATTTAACCAATTATATTAATAAATAAATAAATAAATAAATAAATAAATAA 19752

QY 548 TTTACTGCTTCAAAATTAACATTTGAGTTCAAGTTTGAATACTAAACATAGCATTAATATGAA 607
DB 19751 CTTATATCATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19692

QY 608 TAATGATGGAATAATATATATTCCTTTGAAACTGATTGATTAATATATTCCTCCCTCTT 667
DB 19691 ACAACACATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19633

QY 668 AGAAACAGTCAAAAGCCACTTCAAAACAAAGTTTCAAAATAAAGGAGGTAGCAAGTTAGGC 727
DB 19632 AAAAAACCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19573

QY 728 GATGATTAATATTTTCTTGGCTTGTGTTATACCAATTTGGCCAGGCGCTTTATAGGACTC 787
DB 19572 ATTAATAATACTAATTTTCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19513

QY 788 CCATAAGCATTTTGAAGATGCAATATCAATAATAGTGTATGTCCTCTCAATGAGGCAT 847
DB 19512 AATAATCCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19453

QY 848 TTTTAAATGTTTACATCTATTGTCAGGCTTCAGGTTATGATATGTTATGAAATAAGCT 907

XX 07-DEC-2001 (first entry)
DT Human breast cancer expressed polynucleotide 7667.
DE Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX WO200151628-A2.
XX 19-JUL-2001.
XX 10-JAN-2001; 2001WO-US000798.
XX 14-JAN-2000; 2000US-0176077P.
XX 14-MAR-2000; 2000US-0189167P.
XX 24-MAR-2000; 2000US-0192099P.
XX 29-MAR-2000; 2000US-0193480P.
XX 15-MAY-2000; 2000US-0205230P.
XX 09-JUN-2000; 2000US-0211315P.
XX 25-JUL-2000; 2000US-0220534P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
FA Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX New peptide useful as a marker for the diagnosis of breast cancer.
XX Claim 1; Page 1378; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity
XX Sequence 883 BP; 322 A; 18 C; 23 G; 396 T; 0 U; 124 Other;
SQ

Query Match 3.0%; Score 48.8; DB 4; Length 883;
Best Local Similarity 38.3%; Pred. No. 0.45;
Matches 206; Conservative 0; Mismatches 332; Indels 0; Gaps 0;

QY 440 TTATAATTAATGATCTGTAATAAATCAATACACAGGATAAGATCTCAAAATGAGTGTGACT 499
DB TTTTATATAAANAATTTTTTTTAAATATAAATANAATGANGTTTAANTTATT 319

QY 500 CCCTTAAAGTAGATTAAGTGTGCAATCTTTGTTTCTTAAATATGATTTTACTGCTTGA 559
DB TTTATTTAATTAANAATAAATTTTNTNTNANTTTTAAATTTNTNTNNTNNTAATAA 379

QY 560 AATTACATTTGAGTTGAGTTAGAACTAACAATAGCAATTAATATGAATATGATCGAA 619
DB ATTATTTTNNAAATTTATTTTNTNTNTNTNTTTTAAATTTATAGAATAAANAATAA 439

QY 620 AATTATTATCTTTGAAACTGATGATAAATATATATCCCTCTTCTTAAAGACAGTCAA 679
DB ATTTTNTAATNTNNAANNAANNAANNTNTNTTNAATTAAGTTTAAANAANNTTN 499

QY 680 AGCCACTTCAACAAGTTTCAATATAAGGAGGTAGCAAGTTAGGCGATGATTATAT 739
DB AAAAAATTTTTTTTTTTTNTTNTTNGAGAAAANAANAATAAATTTTTTATAANTATAA 559

QY 740 TTTCTTGGCTTGTATACCAATGGCCAGGCTTTTATAAGGACTCCCAAGCAATTT 799
DB AAAAAANNTTTTTTTTAAATATNTNTNATAAANAANNTNTNTTTTATTTTAAANNTTT 619

QY 800 TGAAGATGGCAATATCAAAATAAGTGTATCTCTCAAAATGAGGCAATTTTAAATCTTA 859
DB TTTATNNAANAANAANAANAANANTTGNTTAAATTTNARATTTTTTTTTTATATAAAA 679

QY 860 CAATCTATTTGGACGCTCAGGTTATGATGTTATGAAAAAATAGCTTCATTTCTT 919
DB AAATNNAANTTATTAANTTTTTTAAAAAATTTTTTAAAAAANAATTTTGGTTAATAANT 739

QY 920 ATAGCTACATCTATTATTCCTTTTAGAAAACAAGATAACAATAAGCTTTTAAATAGTT 977
DB TAAAAAANNTTAAATTTTTTNNNTTTTTTANAANAATNTANTANTANTATAGTT 797

RESULT 14
ABZ10246/c
ID ABZ10246 standard; DNA; 8056 BP.
XX AC ABZ10246;
XX 16-JAN-2003 (first entry)
XX Haematopoietic cell proliferation disorder related DNA sequence #386.
XX Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX Homo sapiens.
XX WO20027272-A2.
XX 03-OCT-2002.
XX 26-MAR-2002; 2002WO-EP003401.
XX 26-MAR-2001; 2001US-0278333P.
XX (EPIG-) EPIGENOMICS AG.
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwöpe I, Ziebarth H;
XX WPI; 2003-018942/01.

DR Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX Claim 28; SEQ ID NO 386; 117pp; English.

XX The present invention describes a method for detecting and
XX differentiating between haematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118
XX represent specifically claimed nucleotide sequences from the present
XX invention. Oligonucleotides from the present invention can be used: for
XX differentiating between healthy haematopoietic cells and proliferative
XX disorder haematopoietic cells; for differentiating between acute
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
XX determining the cytosine methylation state and/or single nucleotide
XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder
XX related sequences and their complements; and as primers for the
XX amplification of haematopoietic cell proliferation disorder related DNA
XX sequences. The nucleotide sequences from the present invention can also
XX be used for detecting a predisposition to, differentiation between
XX subclasses, diagnosis, prognosis, treatment and/or monitoring of
XX haematopoietic cell proliferative disorders. The present method enables a
XX highly specific classification of haematopoietic cell proliferative

CC disorders allowing for improved and informed treatment of patients

XX SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 3.0%; Score 48.8; DB 7; Length 8056;
Best Local Similarity 44.0%; Pred. No. 0.72;
Matches 297; Conservative 0; Mismatches 374; Indels 4; Gaps 2;

QY 438 ACTTAAATAGATGCTGAAATAACATACCAAGGATAAGATGTCAAATGAGTGA 497
Db |||||
QY 2148 ACTTTTAAATAAATAAATAATTAATAATTAATAATTAATAATTAATTAAT 2089
Db |||||
QY 498 CTCCTTAAAGTAGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 557
Db |||||
QY 2088 AATATTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTT 2029
Db |||||
QY 558 GAAATACATTTGAGTGAAGTTTGAAGAA--CTAACATAGCATTAATGATGATCA 614
Db |||||
QY 2028 TATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTT 1969
Db |||||
QY 615 TCGAAATATTTATCTTTGAAACATGATGATATATATATATATATATATATAT 674
Db |||||
QY 1968 TTTTAAATATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1909
Db |||||
QY 675 GTCAAAAGCCACTTCAAAACAGTTTCAAAATAAAGGAGGTAGCAAGTTAGCGATGGAT 734
Db |||||
QY 1908 ATTTTATTTTATATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1849
Db |||||
QY 735 TATATTTTCTGGCTTGTGTATACCCATGCGCGGCTTTATAAGGACTCCCAAG 794
Db |||||
QY 1848 TTTTATTTTATTTTAAACAAATTTTATTTTATTTTATTTTAAATTAATAAT 1789
Db |||||
QY 795 CATTTTGAAGATGGAATATCAATATCAATATCAATATCAATATCAATATCAAT 854
Db |||||
QY 1788 TAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1729
Db |||||
QY 855 TGTACATCTATTTGGAGCTCAGGTTATGATATGTTTATGAAATTAAGCTTCATTT 914
Db |||||
QY 1728 TATAATATATATATTTA-TATTAATTAATTTTATATATTTTAAATAAATAA 1670
Db |||||
QY 915 TTCTTATAGTACATCTCTATTTCTCTTTTAAACCAAGATAAATAAGTTTAA 974
Db |||||
QY 1669 ATCATATTAATAATAATTTTATTTTATTTTATTTTATTTTATTTTATTT 1610
Db |||||
QY 975 GTTGCCATCTTAGCATTTATCAGGCTATGAAACCAATATTTGATCTCTGATTAAT 1034
Db |||||
QY 1609 ATATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 1550
Db |||||
QY 1035 TTCTGATCTTACTAGCTATGGGAATTTAGAACTGGCACACCTGACATTAAGTGG 1094
Db |||||
QY 1549 AATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1490
Db |||||
QY 1095 AATGTTAGGATTTT 1109
Db |||||
QY 1489 ATAATTAATTTT 1475
Db |||||

RESULT 15

ABL32891/c
ID ABL32891 standard; DNA; 6131 BP.

XX ABL32891;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 864.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; anianaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.

XX Claim 1; SEQ ID NO 864; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention

XX SQ Sequence 6131 BP; 1970 A; 39 C; 1013 G; 3109 T; 0 U; 0 Other;

Query Match 3.0%; Score 48.6; DB 6; Length 6131;

Best Local Similarity 48.1%; Pred. No. 0.75;

Matches 169; Conservative 0; Mismatches 179; Indels 3; Gaps 1;

QY 888 ATGTTTATGAAATAGCTTCATTTCTTTATAGTACATCTATTTCCCTTTTAG 947
Db |||||
QY 718 ATTATCATTAATAATATTTTACTTTTCTTAAATAATAACCAAAATATATATA 659
Db |||||
QY 948 AACCAAGATAACATAAGTTTAAATAGTTGCCATCTTACGATTTATCAGGCTATGA 1007
Db |||||
QY 658 AACATTTAAATAAATAATAATAATATACATAATCAATATAAAACCAATCATATAATTA 599
Db |||||
QY 1008 AACCAATTTGAATCTCTGATAAATAATTTCTGATGTTACTAGTATGGGAATTAGAAC 1067
Db |||||
QY 598 AATTTATTTCAATATATATTTTACTTTCTTCTTAAATAAATTCATTTATCATCAAC 539
Db |||||
QY 1068 TGGCAACACCTGACATTTACTTAAGTGGAAATGTTAGGATTTTTCGGCATCGATTTAGA 1127
Db |||||
QY 538 TAATATTATCATTAACCTTAAATTTCTTATAAACAATAATTTACTTATATATATCAATAATA 479
Db |||||
QY 1128 ATCTCTAAATTTTAAACATTTCCCTGTTAAATGACTAAGTTTTCCTTTATCAATATGATTT 1187
Db |||||
QY 478 AA---TATTATATAAATTTTAAACCCCAATTAATAATACAAATTTATACAAATTAATTA 422
Db |||||
QY 1188 CTGAAGGCCAATATCATACCATTAACATATGAAAGCTTTTAAATTCCTTAAAA 1238
Db |||||
QY 421 TTCAATAAATAATTTTCTTCTTAAATTTATATAAACAATATATTTCTTAAAA 371
Db |||||

Search completed: April 25, 2004, 08:46:03

Job time : 434.176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:54:27 ; Search time 4263.17 Seconds
(without alignments)
16460.143 Million cell updates/sec

Title: US-10-051-835-15

Perfect score: 1619

Sequence: 1 ttaattcacatatctatgt.....atttatttaaaatttttaaaa 1619

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hg_hum.*
31: em_hg_inv.*
32: em_hg_other.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rtd.*
36: em_hg_nam.*
37: em_hg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1417.4	87.5	5900	9	HSN805990	BX537884 Homo sapi
2	1415	87.4	157042	9	AC024061	AC024061 Homo sapi
3	1414.2	87.4	2150	9	BC058026	BC058026 Homo sapi
4	1412.6	87.3	5292	9	HSN808902	BX548751 Homo sapi
5	1412.6	87.3	6104	9	HSN808720	BX548751 Homo sapi
6	1411.8	87.2	176626	9	AC066614	AC066614 Homo sapi
7	1410.2	87.1	2455	9	AK123805	AK123805 Homo sapi
8	1408.6	87.0	3493	9	AK125130	AK125130 Homo sapi
9	475	29.3	580	6	AX332151	AX332151 Sequence
10	475	29.3	580	6	AX334874	AX334874 Sequence
11	475	29.3	580	6	AX335287	AX335287 Sequence
12	475	29.3	580	6	AX335859	AX335859 Sequence
13	118.2	7.3	231508	10	AC111087	AC111087 Mus muscu
14	118.2	7.3	256138	2	AC116714	AC116714 Mus muscu
15	96.2	5.9	202314	2	AC132694	AC132694 Rattus no
16	96.2	5.9	271288	2	AC105817	AC105817 Rattus no
17	68.8	4.2	250029	3	AE014839	AE014839 Plasmodiu
18	66.6	4.1	1141	6	AX083744	AX083744 Sequence
19	62.6	3.9	111861	9	AC069435	AC069435 Homo sapi
20	62.6	3.9	148075	9	AC110014	AC110014 Homo sapi
21	60.4	3.7	253001	3	AE014834	AE014834 Plasmodiu
22	60	3.7	145992	9	AC134919	AC134919 Homo sapi
23	59.8	3.7	7218	6	I66494	I66494 Sequence 14
24	59.6	3.7	151900	9	AC107419	AC107419 Homo sapi
25	58.8	3.6	212046	10	BX005167	BX005167 Mouse DNA
26	58.8	3.6	316399	2	AC128737	AC128737 Mus muscu
27	58.4	3.6	15421	3	PFCOMPIRA	X95275 P.falciparu
28	58.4	3.6	175559	2	AC145101	AC145101 Homo sapi
29	56.6	3.5	192329	2	AC005505	AC005505 Plasmodiu
30	56.6	3.5	250713	3	AE014850	AE014850 Plasmodiu
31	56.6	3.5	256172	2	AC005139	AC005139 Plasmodiu
32	56.4	3.5	213591	2	AC141759	AC141759 Apis mell
33	56.2	3.5	172816	9	AC093899	AC093899 Homo sapi
34	55.8	3.4	49306	3	AC115606	AC115606 Dictyoste
35	55.6	3.4	169494	9	AL450083	AL450083 Human DNA
36	55.4	3.4	14001	3	PFCOMPIRB	X95276 P.falciparu
37	55.2	3.4	1141	6	AX083744	AX083744 Sequence
38	55.2	3.4	186165	9	AC112187	AC112187 Homo sapi
39	55	3.4	257757	3	AE014837	AE014837 Plasmodiu
40	54.8	3.4	133877	2	AC120883	AC120883 Homo sapi
41	54.6	3.4	14867	3	AE001398	AE001398 Plasmodiu
42	54.4	3.4	194074	9	AC090095	AC090095 Homo sapi
43	54.4	3.4	199810	9	AC018443	AC018443 Homo sapi
44	53.8	3.3	101739	9	AL591400	AL591400 Human DNA
45	53.8	3.3	122364	9	AC007324	AC007324 Homo sapi

ALIGNMENTS

RESULT 1
HSN805990
LOCUS HSN805990 5900 bp mRNA linear PRI 17-JUN-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZ686F07114 (from clone DKFZ686F07114);
complete cds.
ACCESSION BX537884
VERSION BX537884.1 GI:31873891
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5900)
Koehler,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.

Pred. No. is the number of results predicted by chance to have a

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (17-JUN-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center) at the
Heinrich-Heine-University, Duesseeldorf/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp686F07114) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
Location/Qualifiers

FEATURES
source

1..5900
/organism="Homo sapiens"
/mcl_type="mRNA"
/db_xref="taxon:9606"
/map="15q21.3"
/clone="DKFZp686F07114"
/tissue_type="human fetal kidney"
/clone_lib="686 (synonym: hlcc3). Vector pSPORT1_sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="fetal"
1..5900
/gene="DKFZp686F07114"
43..3351
/gene="DKFZp686F07114"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD97880.1"
/db_xref="GI:31873892"
/tranlation="MRTSLQAVALLWGQKAPPHSITAITDDORTIVTQSGEQCPCLW
NLSHELKLSAKELLFGHSASVTCCLARDESKOPYLSVAASGEMCVNVTQSGCVCEV
ATLPYRHTALCYTHCSFRMTGEGWLLCCGEYQVLIIDAKTLAVHSPSSQPPDWIN
CMCIYHMRLLQESLVLVSVAGGKVDLSLLTEVSRNGQFPAGGEVLAARHLIWTGG
RFTCYTIVLQVSSCKWYDYCDKSWLLTSVRSQVQKQDYVEKSKFLESLNCTPI
HSYIYOLLNGLSKSIYPADGRVLYKETIYLHLCLSTSVQNEQSRPFVNGYNNERKE
PFYKVLFGSGVGRITLWH; PDVPSVFGSDPREIPVATLTQDNFHDWTMSQII
DYFSGLQDGAGTAVWSSEYIPLSDKLI CGCEDGTIIITQALNAAKARLLEGGSLVKD
SPHKVLKGHQHSVTSLLYPHGLSSKLDQSMLSGLDSCVILWDI FTBELHKKFPLE
AGPVLSLLSPFKLREGEOIICVCGDHSVALLHLEGSKLLHASKHLFPVPMIKWH
PVENLIVGACDSVYIWEIETGTLEHRTGERARILINCCDDPOLVKSVLPIASLT
KHKSIORSVSOLGCPGQVLESSCKVDAKPCFPFPFVLPVKTKMNSVNFHIL
LFPLENLVEILLPTPSLSDVSSSFGVEGVLRAKSTVEKKTITLRKSTACGFLP
ALAKPIYESLAQDNTI FSEENDGI KRQKMKI SKQMKPKPKVKVDASLTIDTAKFL
LSGLLPWGVKDLIDYLCKHLNLIKQPI SLGTSINEDNFSMLPQWDLCSNGMIK
YSCVPLKGRVLDLSDKVTATLPNOVGI PRGLNCCDSRESPTIVLLSRLLVANK
VNWPLELACVQSGSFRMESIHNKVRGAGNDILANSSPYSLRNGKSHVPEADLSLL
KLISWRDQSVQVTEATQAVLLFAVQHKSLGKI PYNQSPVSNANGENCKMQLPK
LEWTELEQLQVNRNLTFLQIPVSPVFKDSSNSNANFQVDEMDPRCALEESSESFGPRR
HHSWIAKVCPCKVS"
polya_signal
5867..5872
/gene="DKFZp686F07114"
polya_site
5882
/gene="DKFZp686F07114"

ORIGIN

Query Match 87.5%; Score 1417.4; DB 9; Length 5900;
Best Local Similarity 96.1%; Pred. No. 1.3e-262;
Matches 1556; Conservative 0; Mismatches 50; Indels 13; Gaps 11;

QY 1 TTAATCTACATATCTTATGTATATAAAATATTGCCTAATCCATGAAAAAGGATATAAG 60
Db 4289 TTAATCTACATATCTTATGTATATAAAATATTGCCTAATCCATGAAAAAGGATATAAG 4348

QY 61 TATTAAATATGTGATATATACGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATAGGAGAGTTCAA 120
Db 4349 TATTAAATATGTGATATATAGCTATATCTATCTATCTATGTATCTTATAGGGAAGTTCAA 4408

QY 121 GTCACCTTCAAATTGAAGAACATATCTTGAGCATAGGACGACCTCAGGTCTTATGCTGG 180
pb 4408 CTCACCTTCAAATTGAAGAACATATCTTGAGCATAGGACGACCTCAGGTCTTATGCTGG 4468

```
QY 1361 ATGCTCTCTTAATCCATACGCAAGTGTGTTTATGACACAAATTCAGTCTGGTTA 1320
Db 5545 AIGCTCTCTTAATCCATACGCAAGTGTGTTTATGACACAAATTCAGTCT-GTTA 5603
QY 1321 AAAATGAAATCTTTATATGCTGGGTGCCACATATTCAGTAAATTCGTATGAGA 1380
Db 5604 AAAATGAAATCTTTATATGCT-GGTGTCCACATATTCAGTAAATTCGTATGAGA 5662
QY 1391 GGACTTGAATAGCAAAATGGCCACACAGTTAACTGGATAGACACAGTGTGTGATCA 1440
Db 5663 GGACTTGAATAGCAAAATGG-CCACACAGTTAACTGGATAGACACAGTGTGTGATCA 5721
QY 1441 TAACCACTGTGTACTACACCCAGAACTCAAATTTGCTTTCTCTCGATGAGATATGGG 1500
Db 5722 TAACCACTGTGTACTACACCCAGAACTCAAATTTGCT-TTCTCTCGATGAGATAT-OG 5779
QY 1501 GTGTCTTTTGTAGCTGTAGGGCTAGGTCACCAAGTGAAGTGAATATATACCAAAAT 1560
Db 5780 GTGTCTTTTGTAGCTGTAGGGCTAGGTCACCAAGTGAAGTGAATATATAGC-AAAT 5835
QY 1561 GTGTTTGTATCCAGAGTCTTCTCTGTCAATGTAAATAAAAAATTTATTTAAAAATTTAAA 1619
Db 5836 GTGTTTGTATCCAGAGTCTTCTCTGTCAATGTAAATAAAAAATTTATTTAAAAATTTAAA 5894

RESULT 2
AC024061/c 157042 bp DNA linear PRI 10-JUL-2001
LOCUS Homo sapiens chromosome 15 clone RP11-232J12 map 15q21.3, complete
DEFINITION
ACCESSION AC024061
VERSION HTG.
KEYWORDS AC024061.7 GI:14647263
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157042)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 157042)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,S.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (22-FEB-2001) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 157042)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submission
Submitted (10-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Jul 10, 2001 this sequence version replaced gi:13162487.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: data AC066614 [Drafting center UMWSC] and AC066611
[Drafting center UMWSC] were added for finishing.
Location/Qualifiers

FEATURES
QY 121 ATGCTCTCTTAATCCATACGCAAGTGTGTTTATGACACAAATTCAGTCTGGTTA 1320
Db 5545 AIGCTCTCTTAATCCATACGCAAGTGTGTTTATGACACAAATTCAGTCT-GTTA 5603
QY 1321 AAAATGAAATCTTTATATGCTGGGTGCCACATATTCAGTAAATTCGTATGAGA 1380
Db 5604 AAAATGAAATCTTTATATGCT-GGTGTCCACATATTCAGTAAATTCGTATGAGA 5662
QY 1391 GGACTTGAATAGCAAAATGGCCACACAGTTAACTGGATAGACACAGTGTGTGATCA 1440
Db 5663 GGACTTGAATAGCAAAATGG-CCACACAGTTAACTGGATAGACACAGTGTGTGATCA 5721
QY 1441 TAACCACTGTGTACTACACCCAGAACTCAAATTTGCTTTCTCTCGATGAGATATGGG 1500
Db 5722 TAACCACTGTGTACTACACCCAGAACTCAAATTTGCT-TTCTCTCGATGAGATAT-OG 5779
QY 1501 GTGTCTTTTGTAGCTGTAGGGCTAGGTCACCAAGTGAAGTGAATATATACCAAAAT 1560
Db 5780 GTGTCTTTTGTAGCTGTAGGGCTAGGTCACCAAGTGAAGTGAATATATAGC-AAAT 5835
QY 1561 GTGTTTGTATCCAGAGTCTTCTCTGTCAATGTAAATAAAAAATTTATTTAAAAATTTAAA 1619
Db 5836 GTGTTTGTATCCAGAGTCTTCTCTGTCAATGTAAATAAAAAATTTATTTAAAAATTTAAA 5894

Query Match 87.4%; Score 1415; DB 9; Length 157042;
Best Local Similarity 96.2%; Pred. No. 26-262;
Matches 1553; Conservative 0; Mismatches 49; Indels 13; Gaps 11;

QY 1 TTAATCTCATATCTTATATATATATATATATATATATATATATATATATATATATAT 60
Db 17095 TTAATCTCATATCTTATATATATATATATATATATATATATATATATATATATAT 17036
QY 61 TATTAATATGTGATATATAGTATATATATATATATATATATATATATATATATATATAT 120
Db 17035 TATTAATATGTGATATATAGTATATATATATATATATATATATATATATATATATATAT 16976
QY 121 GTCACCTTCAATTTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGTCTCTATGGTGG 180
Db 16975 GTCACCTTCAATTTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGTCTCTATGGTGG 16916
QY 181 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 240
Db 16915 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 16856
QY 241 ATAAAGCATTAATATATATATATATATATATATATATATATATATATATATATATATAT 300
Db 16855 ATAAAGCATTAATATATATATATATATATATATATATATATATATATATATATATATAT 16796
QY 301 AAGTCACTTCAATTAAGAAACATATTTTTCAGCATGGAGCAGCAGCTCTATCTCTG 360
Db 16795 AAGTCACTTCAATTAAGAAACATATTTTTCAGCATGGAGCAGCAGCTCTATCTCTG 16736
QY 361 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 420
Db 16735 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 16676
QY 421 GGCTGCACTTAAAGAAACATATATATATATATATATATATATATATATATATATATATAT 480
Db 16675 GGCTGCACTTAAAGAAACATATATATATATATATATATATATATATATATATATATATAT 16616
QY 481 ATGTCAAAATGAGTGTGATCTCTCTTAAAGTAGATTAAGTGTGATCTCTCTCTCTCTCTCT 540
Db 16615 ATGTCAAAATGAGTGTGATCTCTCTTAAAGTAGATTAAGTGTGATCTCTCTCTCTCTCTCT 16556
QY 541 ATATGATTTTACTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
Db 16555 ATATGATTTTACTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16496
```

source

```
1. .157042
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.3"
/clone="RP11-232J12"
/clone_lib="RPCI Human BAC library 11"
/note="This clone overlaps RP11-808 AC066614 and
RP11-348P16 AC066611. Data from overlapping BACs were
combined and the consensus sequence determined from
RP11-232J12 to the extent possible."
misc_feature
1. .26818
/note="overlap with RP11-808 AC066614"
unsure
8723
/note="low quality data."
31605..31610
/note="low quality data."
34826
/note="low quality data."
42795..42800
/note="low quality data."
63513..63514
/note="low quality data."
124490..124492
/note="low quality data."
misc_feature
143570..157042
/note="overlap with RP11-348P16 AC066611."
```

ORIGIN

```
Query Match 87.4%; Score 1415; DB 9; Length 157042;
Best Local Similarity 96.2%; Pred. No. 26-262;
Matches 1553; Conservative 0; Mismatches 49; Indels 13; Gaps 11;

QY 1 TTAATCTCATATCTTATATATATATATATATATATATATATATATATATATATAT 60
Db 17095 TTAATCTCATATCTTATATATATATATATATATATATATATATATATATATATAT 17036
QY 61 TATTAATATGTGATATATAGTATATATATATATATATATATATATATATATATATATAT 120
Db 17035 TATTAATATGTGATATATAGTATATATATATATATATATATATATATATATATATATAT 16976
QY 121 GTCACCTTCAATTTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGTCTCTATGGTGG 180
Db 16975 GTCACCTTCAATTTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGTCTCTATGGTGG 16916
QY 181 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 240
Db 16915 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 16856
QY 241 ATAAAGCATTAATATATATATATATATATATATATATATATATATATATATATATATAT 300
Db 16855 ATAAAGCATTAATATATATATATATATATATATATATATATATATATATATATATATAT 16796
QY 301 AAGTCACTTCAATTAAGAAACATATTTTTCAGCATGGAGCAGCAGCTCTATCTCTG 360
Db 16795 AAGTCACTTCAATTAAGAAACATATTTTTCAGCATGGAGCAGCAGCTCTATCTCTG 16736
QY 361 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 420
Db 16735 GATGCACTGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 16676
QY 421 GGCTGCACTTAAAGAAACATATATATATATATATATATATATATATATATATATATATAT 480
Db 16675 GGCTGCACTTAAAGAAACATATATATATATATATATATATATATATATATATATATATAT 16616
QY 481 ATGTCAAAATGAGTGTGATCTCTCTTAAAGTAGATTAAGTGTGATCTCTCTCTCTCTCTCT 540
Db 16615 ATGTCAAAATGAGTGTGATCTCTCTTAAAGTAGATTAAGTGTGATCTCTCTCTCTCTCTCT 16556
QY 541 ATATGATTTTACTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
Db 16555 ATATGATTTTACTGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16496
```

QY	601	ATATGAATATGATGGAATAATTTATATCTCTTTGAAACATGATGATGAATAATATATTTCCCC	660
Db	16495	ATATGAATAT - CATGGAATAATTTATATCTCTTTGAAACATGATGATGAATAATATTTCCCC	16437
QY	661	CTCCTTTAGAAACAGTCAAAAGCCTTCAAAACAGTTTCAAAATTAAGGAAGGTAGCAA	720
Db	16436	CTCCTTTAGAAACAGTCAAAAGCCTTCAAAACAGTTTCAAAATTAAGGAAGGTAGCAA	16377
QY	721	GTAGGCGATGGATATATATTTCTTGGCTTGTGTATACCATGTCAGGCCCTTTATA	780
Db	16376	GTAGGCGATGGATATATATTTCTTGGCTTGTGTATACCATGTCAGGCCCTTTATA	16319
QY	781	AGACTCCCAAGCAATTTTGAAGATGCAATATCAAAATAGTGTATCTCTCAAT	840
Db	16318	AGACTCCCAAGCAATTTTGAAGATGCAATATCAAAATAGTGTATCTCTCAAT	16259
QY	841	GAGGCATTTTAAATGTTCAATCTATTTGGAGCTCAGGTATGATATGTTTATGAAA	900
Db	16258	GAGGCATTTTAAATGTTTAAATCTATTTGGAGCTCAGGTATGATATGTTTATGAAA	16199
QY	901	ATAGCTTCAATTTCTTATAGCTACATCTTATTTCCCTTTTGAACCAAGATAC	960
Db	16198	ATAGCTTCAATTTCTTATAGCTACATCTTATTTCCCTTTTGAACCAAGATAC	16140
QY	961	AATAAGTTTAAATGTTGCGCATTTAGCAATTTATCAGTCTAATGAACCAATTTGAA	1020
Db	16139	AATAAGTTTAAATGTTGCGCATTTAGCAATTTATCAGTCTAATGAACCAATTTGAA	16080
QY	1021	TCCTCATATAATTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTG	1080
Db	16079	TCCTCATATAATTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTG	16020
QY	1081	ACATTACTAAGTGAATGTTAGGATTTTTCGCGCATCGCATGTTAGAACTCTCTAAATTT	1140
Db	16019	ACATTACTAAGTGAATGTTAGGATTTTTCGCGCATCGCATGTTAGAACTCTCTAAATTT	15960
QY	1141	AAACATTCCTGTTAAATGATAGTGTGCTTTTATCAATATCAATCTCGAAGCCATA	1200
Db	15959	AAACATTCCTGTTAAATGATAGTGTGCTTTTATCAATATCAATCTCGAAGCCATA	15900
QY	1201	TCATACCATTAACATGAAAGCTTTTAAATTCCTAAAAATAGTTTTAGAGATTTCAAGCA	1260
Db	15899	TCATACCATTAACATGAAAGCTTTTAAATTCCTAAAAATAGTTTTAGAGATTTCAAGCA	15840
QY	1261	ATGCTCTCTAATATCCATGCAAGTGTGTTTATGACAAATTCACATGCTGTGTTTA	1320
Db	15839	ATGCTCTCTAATATCCATGCAAGTGTGTTTATGACAAATTCACATGCTGTGTTTA	15781
QY	1321	AAATGAAATCTTTATATTGACTGGGTGCCACATATTCAGTAATTTCTGTTATGAGA	1380
Db	15780	AAATGAAATCTTTATATTGACT - GGTGTTCCACATATTTTCAGTAATTTCTGTTATGAGA	15722
QY	1381	GGACTTGAATAGCAATTCGCCACACAGTTAACTGGATAGATCACTGTCGTTGATCA	1440
Db	15721	GGACTTGAATAGCAATTCGCCACACAGTTAACTGGATAGATCACTGTCGTTGATCA	15663
QY	1441	TAACCACTGCTACTACACCCAGAACTCAAAATGCTTTCTCTGATGATGATGCG	1500
Db	15662	TAACCACTGCTACTACACCCAGAACTCAAAATGCTTTCTCTGATGATGATGCG	15605
QY	1501	GTGTCTCTTTGTACGTTATAGGCGCTAGGTCACCAAGTGAAGTGAATATATAGCAAAAT	1560
Db	15604	GTGTCTCTTTGTACGTTATAGGCGCTAGGTCACCAAGTGAAGTGAATATATAGCAAAAT	15549
QY	1561	GTGTTTGTATCCAGAGTCTCTCTGTCATCTGATTAATAAAATTTTATTAATAATTT	1615
Db	15548	GTGTTTGTATCCAGAGTCTCTCTGTCATCTGATTAATAAAATTTTATTAATAATTT	15494

RESULT 3
BC058026
LOCUS

DEFINITION	Homo sapiens cDNA clone IMAGE:4608833, partial cds.
ACCESSION	BC058026
VERSION	BC058026.1 GI:34785600
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 2150) Strausberg, R.L., Reingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Whitting, M., Kettman, M., Madan, A., Young, A.C., Rodriguez, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, F.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
TITLE	2388257 12477332
JOURNAL	2 (bases 1 to 2150)
MEDLINE	Strausberg, R.
PUBMED	Direct Submission
AUTHORS	Submitted (08-SEP-2003): National Institutes of Health, Mammalian
TITLE	Gene Collection (MGC), Cancer Genomics Office, National Cancer
JOURNAL	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
REMARK	USA NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabs-remail.nih.gov Tissue Procurement: CLONTECH cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nhgri.nih.gov Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., McDowell, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDwail, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgenson, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
COMMENT	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate: 48 Row: d Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
FEATURES	Location/Qualifiers 1. 2150 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4608833" /tissue_type="Kidney" /clone_lib="NIH_MGC_75"

RESULT 5	
HSM808720	
LOCUS	6104 bp mRNA linear PRI 30-AUG-2003
DEFINITION	Homo sapiens mRNA; cDNA DKFZP686M05115 (from clone DKFZP686M05115)
ACCESSION	BX648571
VERSION	BX648571.1 GI:34367733
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 6104)
AUTHORS	Bloecher,H., Boecker,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S. The German Human cDNA Consortium Direct Submission Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuerberg, GERMANY
CONSTRM	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
JOURNAL	
TITLE	
COMMENT	

sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp686M05115) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

Location/Qualifiers
1..6104
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="15q21.3"
/clone="DKFZp686M05115"
/tissue_type="human fetal kidney"
/clone_lib="686 (synonym: hicc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stages="fetal"
polyA_signal 5067..6072
polyA_site 6082

ORIGIN

Query Match 87.3%; Score 1412.6; DB 9; Length 6104;
Best Local Similarity 95.9%; Pred. No. 1.1e-261;
Matches 1553; Conservative 0; Mismatches 53; Indels 13; Gaps 11;
1 TTAATCTACATCTTATGTATATAAATATGCTTAATCCATAGAAAAAGGATATAAG 60
4489 TTAATCTACATCTTATGTATATAAATATGCTTAATCCATAGAAAAAGGATATAAG 4548
61 TATTAAATATGATATATAGCCTTAAATATGCTTAATCCATAGAAAAAGGATATAAG 120
4549 TATTAAATATGATATATAGCTATATCTATCTATGATCTAATAGGGAAGTTCAA 4608
121 GTCACTTCAATTAAGAAACATATCTGAGCATAGGAGAGCCCTCAGTCTTATGTTGG 180
4609 GTCACTTCAATTAAGAAACATATCTGAGCATAGGAGAGCCCTCAGTCTTATGTTGG 4668
181 GATGAGTGGCAGAGAGGGGAAATAGAAAAGAGAACTATATAATGAAAAGGGAT 240
4669 GATGAGTGGCAGAGAGGGGAAATAGAAAAGAGAACTATATAATGAAAAGGGAT 4728
241 ATAAAGCAATTAAT 300
4729 ATAAAGCAATTAAT 4788
301 AGTCACCTTCAATTAAGAAACATATCTGAGCATAGGAGAGCCCTCAGTCTTATGTTGG 360
4789 AGTCACCTTCAATTAAGAAACATATCTGAGCATAGGAGAGCCCTCAGTCTTATGTTGG 4848
361 GATGAGTGGCAGAGAGGGGAAATAGAAAAGAGAACTGTGTAATGAAAAGGATGTC 420
4849 GATGAGTGGCAGAGAGGGGAAATAGAAAAGAGAACTGTGTAATGAAAAGGATGTC 4908
421 GGTGACCCCTTAAGAACTTATATATATATATATATATATATATATATATATATATAT 480
4909 GGTGACCCCTTAAGAACTTATATATATATATATATATATATATATATATATATATAT 4968
481 ATGTCAATATGAGTGTGACCTCCCTTAAAGTAGAATTAAGAGTGTGCAATCTTTGTTCTTAA 540
4969 ATGTCAATATGAGTGTGACCTCCCTTAAAGTAGAATTAAGAGTGTGCAATCTTTGTTCTTAA 5028
541 ATATGATTTTACTGCTGAAAT 600
5029 ATATGATTTTACTGCTGAAAT 5088
601 ATATGATTAATGATGAAATATATATCTTGTAAACTGATTCATATAATATATATATATAT 660
5089 ATATGATTAAT -CATGGAATATATATATCTTTGAAACTGATTCATATAATATATATATAT 5147
661 CTCCTTTAGAAAACAGTCAAAAGCCACTTCAAAACAGTTTCAAAATTAAGGAAGGTAGCAA 720

Db 5148 CTCCTTTAGAAAACAGTCAAAAGCCACTTCAAAACAGTTTCAAAATTAAGGAAGGTAGCAA 5207
Qy 721 GTTAGCGATGGATTATATTTCTTGGCTTGTGTATATACCCATTGGCCAGGGCTTTATA 780
Db 5208 GTTAGCGATGGATTATATTTCTT -GCTTGTGTATA -CCAGTTTCAAGGACATTATA 5265
Qy 781 AGGACTCCCAAGACATTTTGAAGATGCAATATCAAAATAGTGTATGTCCTCTCAAAAT 840
Db 5266 AGGACTCCCAAGACATTTTGAAGATGCAATATCAAAATAGTGTATGTCCTCTCAAAAT 5325
Qy 841 GAGGCATTTTATTTTATTTTCAATTTTGGAGCGCTCAGGTATGATATGTTTATGAAAA 900
Db 5326 GAGGCATTTTATTTTATTTTCAATTTTGGAGCGCTCAGGTATGATATGTTTATGAAAA 5385
Qy 901 ATAGCTTCATTTATTTTATAGTACATCTCTATATATCCCTTTTGAAGAACAGATATAC 960
Db 5386 ATAGCTTCATTTATTTTATAGTACATCTCTATATATCCCTTTTGAAGAACAGATATAC 5444
Qy 961 AATAAGTTTAAATAGTTGCAATCTTATAGCTTATCAGGTCTAATGAACCAATATTGAA 1020
Db 5445 AATAAGTTTAAATAGTTGCAATCTTATAGCTTATCAGGTCTAATGAACCAATATTGAA 5504
Qy 1021 TCTGATATAAATTTTCTGATTTAGTATCTAGCTATGGGAAATTAGAACTGGCACAACCTG 1080
Db 5505 TCTGATATAAATTTTCTGATTTAGTATCTAGCTATGGGAAATTAGAACTGGCACAACCTG 5564
Qy 1081 ACATTTACTAGTGAATTTTGTAGGATTTTGGCATCGCATGTTAGAATCTCTTAAAAATTT 1140
Db 5565 ACATTTACTAGTGAATTTTGTAGGATTTTGGCATCGCATGTTAGAATCTCTTAAAAATTT 5624
Qy 1141 AATCATTTCTTTTAAATAGTAAAGTTTGTAGGATTTTCAATATGAATCTTGAAGGCCAATA 1200
Db 5625 AATCATTTCTTTTAAATAGTAAAGTTTGTAGGATTTTCAATATGAATCTTGAAGGCCAATA 5684
Qy 1201 TCATACCATTAATCTAGAAAGCTTTTAAATCTTAAATAATAGTTTGAAGATATTCAAGCA 1260
Db 5685 TCATACCATTAATCTAGAAAGCTTTTAAATCTTAAATAATAGTTTGAAGATATTCAAGCA 5744
Qy 1261 ATGCTCTCTCTTAAATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGTTT 1320
Db 5745 ATGCTCTCTCTTAAATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCT -GTTT 5803
Qy 1321 AATATGAATCTTTATATATGACTGGGTGTCACATATTTCCAGTAAATTTCTGTTATGAGA 1380
Db 5804 AATATGAATCTTTATATATGACT -GGTGTTCACATATTTCCAGTAAATTTCTGTTATGAGA 5862
Qy 1381 GGACTTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTGCATCA 1440
Db 5863 GGACTTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTGCATCA 5921
Qy 1441 TAACCACTTGGTACTACCCAGAAATCTCAAAATTTGCTTTTCTCTCTGATGAGATATGGG 1500
Db 5922 TAACCACTTGGTACTACCCAGAAATCTCAAAATTTGCT -TTTCTCTCTGATGAGATAT -GG 5979
Qy 1501 GTGTCTCTTTTGTACGTCTAGGGCTTAGGGTACCCAAAGTGAAGTGAATATATATAGCAAAAT 1560
Db 5980 GTGTCTCTTTTGTACGTCTAGGGCT -TAGTGAACCAAGTGAATATATATATAGC -AAAT 6035
Qy 1561 GTGTGTGTATCCAGAGTCTTCTGTCATGTAATAAAAAATTTTATTTAAAAATTTAAAA 1619
Db 6036 GTGTGTGTATCCAGAGTCTTCTGTCATGTAATAAAAAATTTTATTTAAAAATTTAAAA 6094

RESULT 6

AC066614/c
LOCUS Homo sapiens chromosome 15 clone RF11-808 map 15q21.3, complete
DEFINITION AC066614
ACCESSION AC066614
VERSION AC066614.8
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens


```
QY 961 AATAAGCTTTTAAAGTGGCCTAGCTATAGCATTTATCAGGCTTAATGAAACCAATATTGAA 1020
Db 1818 AATAGTGTAAATAGTGGCCTAGCTATAGCATTTATCAGGCTTAATGAAACCAATATTGAA 1877
QY 1021 TCTCTGATAAATAATTTCTGATGTTACTAGCTATAGGAAATAGAACTGGCAACACCTTG 1080
Db 1878 TCTCTGATAAATAATTTCTGATGTTACTAGCTATAGGAAATAGAACTGGCAACACCTTG 1937
QY 1081 ACATTACTAGTGAATGTTAGGATTTTGGCATCGCATGTTAGATCTCTAAATTT 1140
Db 1938 ACATTACTAGTGAATGTTAGGATTTTGGCATCGCATGTTAGATCTCTAAATTT 1997
QY 1141 AAACATTCCTGTTAAATGACTAAGGTTTCTTTTATCAATATGAATTTCTGAAGGCAATA 1200
Db 1998 AAACATTCCTGTTAAATGACTAAGGTTTCTTTTATCAATATGAATTTCTGAAGGCAATA 2057
QY 1201 TCATACCATTAAGTGAAGCTTTTAAATTCCTAAATAGTTTGAAGATTTCAAGCA 1260
Db 2058 TCATACCATTAAGTGAAGCTTTTAAATTCCTAAATAGTTTGAAGATTTCAAGCA 2117
QY 1261 ATGCTCTCTTAATCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGTTT 1320
Db 2118 ATGCTCTCTTAATCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCT-GTTT 2176
QY 1321 AATAGCAATCTTATATGACTGGGTGCTCCACATATTCAGTATATTTCTGTTATGAA 1380
Db 2177 AATAGCAATCTTATATGACT-GGTGTCCACATATTCAGTATATTTCTGTTATGAA 2235
QY 1381 GGACTTTGAATAGCAAAATGCCACACAGTTAACTGGATAGATCACCTACGTTGGTATCA 1440
Db 2236 GGACTTTGAATAGCAAAATGG-CCACACAGTTAACTGGATAGACCACTAGTGGTATCA 2294
QY 1441 TAACCACTTGGTACTACCCAGAACTCAAAATTTCTTTCTCTGATGATGATGG 1500
Db 2295 TAACCACTTGGTACTACCCAGAACTCAAAATTTCT-TTCTCTGATGATGAT-GG 2352
QY 1501 GTGTCCTTTGTACGCTTAGGCGCTAGGCTACCCAGTACCGATGATGATATTAAGCAAA 1560
Db 2353 GTGTCCTTTGTACGCTTAGGCGCTAGGCGCT--TAGGTACCACTAGGATGATATTAAG 2408
QY 1561 GTGTTGTATCCAGAGCTTCTCTGTCATGTTAATAAAAAATTTATTT 1607
Db 2409 GTGTTGTATCCAGAGCTTCTCTGTCATGTTAATAAAAAATTTATTT 2455

RESULT 8
LOCUS AK125130 3493 bp mRNA linear PRI 09-SEP-2003
DEFINITION Homo sapiens cDNA FLJ43140 fis, clone CTONG3007528.
ACCESSION AK125130
VERSION AK125130.1 GI:34531119
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
Kamihara,K., Katsura,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
JOURNAL Unpublished
TITLE (bases 1 to 3493)
REFERENCE 2
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
```

```
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomise@kri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. 3493
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CTONG3007528"
/tissue_type="tongue, tumor tissue"
/clone_lib="CTONG3"
/notes="cloning vector: pME18SFL3"
177..962
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAG86058.1"
/db_xref="GI:34531120"
/translation="MLPGRDLNSGMKDYSGVNLSEKVLDSLDDKVTATLPNOVGIP
RGLNWCDSRESQTIYLLSRPLVNKLWNPELELACRVGSGFRMESHNKRGAGN
DIILNMSFYSLCRNGKNSHVPADLSLLKLISWDRSQSVQVTEAIOAVLLAEVQQHM
KSLGKIPVNSQPVSMAGNCKMQLPKLEMTTELELOCVTEFLQTPVSPVKHDS
NSNSANFQVEDMFDRCALEESESPEPRHSHWIAKVCPCVS"
Query Match 87.08; Score 1408.6; DB 9; Length 3493;
Best Local Similarity 96.28; Pred. No. 7.2e-261;
Matches 1546; Conservative 0; Mismatches 48; Indels 13; Gaps 11;
QY 1 TTAATCTACATATCTTATGATATATAATATTGCTTAATCCATAGAAAAAGGATATAAG 60
Db 1900 TTAATCTACATATCTTATGATATATAATATTGCTTAATCCATAGAAAAAGGATATAAG 1959
QY 61 TATTAATATCTGATATATAGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Db 1960 TATTAATATCTGATATATAGCTATATCTATCTATCTATCTATCTATCTATCTATCTAT 2019
QY 121 GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGGTCCTATGGTGG 180
Db 2020 GTCACTTCAATTGAAGAAACATATCTCTGAGCATAGGAGCAGCTCAGGTCCTATGGTGG 2079
QY 181 GATCAGTGGACAGCAGGCGGGAATTAGAAAAGAGAACTATATTAATTTGAAAAGGGAT 240
Db 2080 GATCAGTGGACAGCAGGCGGGAATTAGAAAAGAGAACTATATTAATTTGAAAAGGGAT 2139
QY 241 ATAAAGCATTAATATATATATATAGCTATATCTATGATGATCTTAACAGAGAAGTTTC 300
Db 2140 ATAAAGCATTAATATATATATATAGCTATATCTATGATGATCTTAACAGAGAAGTTTC 2199
QY 301 AAGTCATCTCAATTAAGAAACATTTTGGAGCATGGGACAGCTCAGGTCCTATGGTGG 360
Db 2200 AAGTCATCTCAATTAAGAAACATTTTGGAGCATGGGACAGCTCAGGTCCTATGGTGG 2259
QY 361 GATCAGTGGACAGCAGTGGGGAATTAGAAAAGAGAACTGTTGTAATTTGAAAATGACGTG 420
Db 2260 GATCAGTGGACAGCAGTGGGGAATTAGAAAAGAGAACTGTTGTAATTTGAAAATGACGTG 2319
QY 421 GGCTGCAACCTTAAGGAACCTTATTAATTAATGATGATCTGTAATAACATACAGGATTAAG 480
Db 2320 GGCTGCAACCTTAAGGAACCTTATTAATTAATGATGATCTGTAATAACATACAGGATTAAG 2379
QY 481 ATGTCAAATGATGCTGACTCCCTTAAAGTAGATTAAGTGTGCAATCTTTGTTTCTCTAAA 540
Db 2380 ATGTCAAATGATGCTGACTCCCTTAAAGTAGATTAAGTGTGCAATCTTTGTTTCTCTAAA 2439
QY 541 ATATGATTTTACTGCTTTGAAATTAACATTTGAGTTGAGTTTGAAGAACTAACATAGCAATTA 600
```

COMMENT

FEATURES
source

CDS

Db 2440 ATATGATTTTACTGCTTGAATTTACATTTGAGTTGAGTTAGAACTAACATAGCATT 2499
QY 601 ATATGAATTAATGATGAAAAATTAATATCTTTGAAACTGATTAATAATATATCCCC 660
Db 2500 ATATGAATAAT-CATGAAAAATTAATATCTTTGAAACTGATTAATAATATATCCCC 2558
QY 661 CTCCTTTAGAACAGTCAAAAGCCACTCAACAGTTTCAAAATAAAGGAAGTAGCAA 720
Db 2559 CTCCTTTAGAACAGTCAAAAGCCACTCAACAGTTTCAAAATAAAGGAAGTAGCAA 2618
QY 721 GTTAGCGATGATTAATTTTCTTGCTTGCTGATATACCAATGGCCAGGCGCTTTATA 780
Db 2619 GTTAGCGATGATTAATTTTCTTGCTTGCTGATATACCAATGGCCAGGCGCTTTATA 2676
QY 781 AGGACTCCCAAGCAATTTGAGAGATGCGCAATATCAATAGTCTATGCTCTCAAT 840
Db 2677 AGGACTCCCAAGCAATTTGAGAGATGCGCAATATCAATAGTCTATGCTCTCAAT 2736
QY 841 GAGGCAATTTTAAATGTTACATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAA 900
Db 2737 GAGGCAATTTTAAATGTTAAATCTATTTGGACGCTCAGGTTATGATATGTTTATGAAA 2796
QY 901 ATAGCTTCATTTATTTCTTATAGCTACATCTTATTTTCCCTTTTAGAAACAAGATAAC 960
Db 2797 ATAGCTTCATTTATTT-TTATAGCTACATCTTATTTTCCCTTTTAGAAACAAGATAAC 2855
QY 961 AATAAGTTTAAATAGTGCCTACTTAGCAATTTATCAGGTTCTAATGAAACCAATATTGAA 1020
Db 2856 AATAAGTTTAAATAGTGCCTACTTAGCAATTTATCAGGTTCTAATGAAACCAATATTGAA 2915
QY 1021 TCTCTGATAAATTTTCTGATGTTACTAGCTATGCGGAATTAGAATGCGGACACCCCTG 1080
Db 2916 TCTCTGATAAATTTTCTGATGTTACTAGCTATGCGGAATTAGAATGCGGACACCCCTG 2975
QY 1081 ACATTTACTAAGTGAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTT 1140
Db 2976 ACATTTACTAAGTGAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTT 3035
QY 1141 AAACATTCCTGTTAAATGATTAAGTTGCTTTTATCAATATGAATTTCTGAAGGCCAATA 1200
Db 3036 AAACATTCCTGTTAAATGATTAAGTTGCTTTTATCAATATGAATTTCTGAAGGCCAATA 3095
QY 1201 TCATACCATTAATGATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGATATTTCAAGCA 1260
Db 3096 TCATACCATTAATGATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGATATTTCAAGCA 3155
QY 1261 ATGCTCTCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGGTTTA 1320
Db 3156 ATGCTCTCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGGTTTA 3214
QY 1321 AAATGAATCTTTATATTTGATGCTGCTCCACATATTCAGTAAATTTCTGTTATGAGA 1380
Db 3215 AAATGAATCTTTATATTTGACT-GGTGTTCCACATATTTTCAGTAAATTTCTGTTATGAGA 3273
QY 1381 GGACTTGAATATAGCAATTTGCCACACACAGTTAACTGGATAGATACGCTAGCTGTGATCA 1440
Db 3274 GGACTTGAATATAGCAATTTGCCACACACAGTTAACTGGATAGATACGCTAGCTGTGATCA 3332
QY 1441 TAACCACTTGTATACACCGAAGTCAAAATTTGCTTTTCTCTGATGAGATATGGG 1500
Db 3333 TAACCACTTGTATACACCGAAGTCAAAATTTGCTTTTCTCTGATGAGATATGGG 3390
QY 1501 GTGCTCTTTTGTACGCTTAGGCGCTAGGCGTACCAGTGAAGTGAATATATAGCAAAAT 1560
Db 3391 GTGCTCTTTTGTACGCTTAGGCGCTAGGCGTACCAGTGAAGTGAATATATAGCAAAAT 3446
QY 1561 GTGTTTGTATCCAGATCTTCTGCTCATTTGTTATTAATAAATTTTATTT 1607
Db 3447 GTGTTTGTATCCAGATCTTCTGCTCATTTGTTATTAATAAATTTTATTT 3493

RESULT 9
AX332151/c

LOCUS AX332151 580 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 2660 from Patent WO0194629.
ACCESSION AX332151
VERSION AX332151.1 GI:18122785
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horgan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 2660 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..580
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 2,6e-81;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
QY 1025 TGTAAATATTTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTGACAT 1084
Db 580 TGTANATATTTTCTGATGTTACTAGCTATGGAAATAGAACTGGCAACCCCTGACAT 521
QY 1085 TACTAAGTGAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTTTAAAC 1144
Db 520 TACTAAGTGAATGTTAGGATTTTTCGGCATCGCATGTTAGATCTCTAAATTTTAAAC 461
QY 1145 ATTCCTGTTAAATGACTAAGTTTGTCTTTTATCAATATGAATTTCTGAAGCCCAATATCAT 1204
Db 460 ATTCCTGTTAAATGACTAAGTTTGTCTTTTATCAATATGAATTTCTGAAGCCCAATATCAT 401
QY 1205 ACATTTAATCATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGATATTCAGCAATGC 1264
Db 400 ACCATTAACATGAAGCTTTTAAATTCCTAAATAATAGTTTATAGATATTCAGCAATGC 341
QY 1265 TCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGTTTAAAAA 1324
Db 340 TCTCCTAATATCCATACGCAAGTGTGTTTATGACACAAATTCACCTAGTCTGTTTAAAAA 282
QY 1325 TGAATCTTTATATGATGCTGGGTGCCACATATTCAGTAATTTCTGTTATGAGGAC 1384
Db 281 TGAATCTTTATATGACT-GGTGTTCCACATATTTTCAGTAATTTCTGTTATGAGGAC 223
QY 1385 TTGAATAGCAAAATTTGCCACACACAGTTAACTGGATAGATCACGTAGTGTGATCATAAC 1444
Db 222 TTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTAGTGTGATCATAAC 164
QY 1445 CACTTGTACTACACCGAAGTCAAAATTTGCTTTTCTCTGATGAGATATGGGTGT 1504
Db 163 CACTTGTACTACACCGAAGTCAAAATTTGCTTTTCTCTGATGAGATATGGGTGT 106
QY 1505 CTTTCTGCTAGCTAGGCGCTAGGCGTACCAGTGAAGTGAATATATAGCAAAATTTGT 1564
Db 105 CTTTCTGCTAGCTAGGCGCTAGGCGTACCAGTGAAGTGAATATATAGCAAAATTTGT 50
QY 1565 TTGATCCAGATCTTCTGCTCATTTGTTATTAATAAATTTTATTTAAAAA 1612
Db 49 TTGATCCAGATCTTCTGCTCATTTGTTATTAATAAATTTTATTTAAAAA 2

RESULT 10
AX334874/c
LOCUS AX334874 580 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 5383 from Patent WO0194629.
ACCESSION AX334874

VERSION AX334874.1 GI:18125593
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Cancer gene determination and therapeutic screening using signature gene sets
JOURNAL Patent: WO 0194629-A 5383 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..580
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 2.6e-81;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
QY 1025 TGATAAATATTTCTGATGTTACTAGCTATGGGAATAGAACTGGCACAACCCCTGACAT 1084
DB 580 TGATANAATATTTCTGATGTTACTAGCTATGGGAATAGAACTGGCACAACCCCTGACAT 521
QY 1085 TACTAAGTGAATATTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTTAAAC 1144
DB 520 TACTAAGTGAATATTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTTAAAC 461
QY 1145 ATTCTCTTAAATGACTAAGGTTTGTCTTATCAATATGAATCTGAAGCCCAATATCAT 1204
DB 460 ATTCTCTTAAATGACTAAGGTTTGTCTTATCAATATGAATCTGAAGCCCAATATCAT 401
QY 1205 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATAGTTTATAGAGATATTCAGCAATGC 1264
DB 400 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATAGTTTATAGAGATATTCAGCAATGC 341
QY 1265 TCTCTTAATATCCATGCAAGTGTGTTATGACACAAATTCAGTCTGTTTAAATA 1324
DB 340 TCTCTTAATATCCATGCAAGTGTGTTATGACACAAATTCAGTCTGTTTAAATA 282
QY 1325 TGAATCTTTATATGACTGGTGTCCACATATTCAGTAAATTCCTGTTATGAGAGGAC 1384
DB 281 TGAATCTTTATATGACTGGTGTCCACATATTCAGTAAATTCCTGTTATGAGAGGAC 223
QY 1385 TTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTATCAATAC 1444
DB 222 TTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTATCAATAC 164
QY 1445 CACTTGGTACTACACCCAGAACTCAAATTTGCTTTTCTCTGATGAGATATGGGTGT 1504
DB 163 CACTTGGTACTACACCCAGAACTCAAATTTGCTTTTCTCTGATGAGATATGGGTGT 106
QY 1505 CCTTTGTACGCTAGGGCTAGGCTACCCAGTGAAGTGAATATATAGCAAAATGTGT 1564
DB 105 CCTTTGTACGCTAGGGCTAGGCTACCCAGTGAAGTGAATATATAGCAAAATGTGT 50
QY 1565 TTGTATCCAGAGTCTTCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
DB 49 TTGTATCCAGAGTCTTCTGTCATTGTAATAAAAAATTTATTTAAAAA 2
RESULT 11
AX335287/c
LOCUS AX335287 580 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 5796 from Patent WO0194629.
ACCESSION AX335287
VERSION AX335287.1 GI:18126006
KEYWORDS Homo sapiens (human)
SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature gene sets
JOURNAL Patent: WO 0194629-A 5796 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..580
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 2.6e-81;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
QY 1025 TGATAAATATTTCTGATGTTACTAGCTATGGGAATAGAACTGGCACAACCCCTGACAT 1084
DB 580 TGATANAATATTTCTGATGTTACTAGCTATGGGAATAGAACTGGCACAACCCCTGACAT 521
QY 1085 TACTAAGTGAATATTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTTAAAC 1144
DB 520 TACTAAGTGAATATTTAGGATTTTCGGCATCGCATGTTAGAACTCTTAAATTTTAAAC 461
QY 1145 ATTCTCTTAAATGACTAAGGTTTGTCTTATCAATATGAATCTGAAGCCCAATATCAT 1204
DB 460 ATTCTCTTAAATGACTAAGGTTTGTCTTATCAATATGAATCTGAAGCCCAATATCAT 401
QY 1205 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATAGTTTATAGAGATATTCAGCAATGC 1264
DB 400 ACCATTAACTATGAAGCTTTTAAATTCCTTAAATAGTTTATAGAGATATTCAGCAATGC 341
QY 1265 TCTCTTAATATCCATGCAAGTGTGTTATGACACAAATTCAGTCTGTTTAAATA 1324
DB 340 TCTCTTAATATCCATGCAAGTGTGTTATGACACAAATTCAGTCTGTTTAAATA 282
QY 1325 TGAATCTTTATATGACTGGTGTCCACATATTCAGTAAATTCCTGTTATGAGAGGAC 1384
DB 281 TGAATCTTTATATGACTGGTGTCCACATATTCAGTAAATTCCTGTTATGAGAGGAC 223
QY 1385 TTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTATCAATAC 1444
DB 222 TTGAATAGCAAAATTTGCCACACAGTTAACTGGATAGATCACGTACGTGGTATCAATAC 164
QY 1445 CACTTGGTACTACACCCAGAACTCAAATTTGCTTTTCTCTGATGAGATATGGGTGT 1504
DB 163 CACTTGGTACTACACCCAGAACTCAAATTTGCTTTTCTCTGATGAGATATGGGTGT 106
QY 1505 CCTTTGTACGCTAGGGCTAGGCTACCCAGTGAAGTGAATATATAGCAAAATGTGT 1564
DB 105 CCTTTGTACGCTAGGGCTAGGCTACCCAGTGAAGTGAATATATAGCAAAATGTGT 50
QY 1565 TTGTATCCAGAGTCTTCTGTCATTGTAATAAAAAATTTATTTAAAAA 1612
DB 49 TTGTATCCAGAGTCTTCTGTCATTGTAATAAAAAATTTATTTAAAAA 2
RESULT 12
AX335859/c
LOCUS AX335859 580 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6368 from Patent WO0194629.
ACCESSION AX335859
VERSION AX335859.1 GI:18126578
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN
Query Match 29.3%; Score 475; DB 6; Length 580;
Best Local Similarity 95.7%; Pred. No. 2.6e-81;
Matches 563; Conservative 0; Mismatches 16; Indels 9; Gaps 7;
1025 TGATAAATATTTCTGATGTTACTAGCTATGGAATTTAGAACTGGCACAACCCCTGACAT 1084
Db TGTATATATTTCTGATGTTACTAGCTATGGAATTTAGAACTGGCACAACCCCTGACAT 521
1085 TACTAAGTGGAAATGTTAGGATTTTCGGCATGCGATGTTAGAAATCTCTAAATTTTAAAC 1144
Db TACTAAGTGGAAATGTTAGGATTTTCGGCATGCGATGTTAGAAATCTCTAAATTTTAAAC 461
1145 ATTCTGTTTAAATGACTAAGGTTTCTTTTCAATATGAATCTTGAAGGCCAATATCAT 1204
Db ATTCTGTTTAAATGACTAAGGTTTCTTTTCAATATGAATCTTGAAGGCCAATATCAT 401
1205 ACCATTAACTATGAAGCTTTTAAATCTTCAAAAATAGTTTATAGATATTCAGCAATGC 1264
Db ACCATTAACTATGAAGCTTTTAAATCTTCAAAAATAGTTTATAGATATTCAGCAATGC 341
1265 TCTCCTAATATCATACGCAAGTGTTTATGACACAAATTCACAGTCTGTTTAAAAA 1324
Db TCTCCTAATATCATACGCAAGTGTTTATGACACAAATTCACAGTCTGTTTAAAAA 282
1325 TGAATATCTTTATTTGACTTCCACATATTTCCAGTAATTTCTGTTATGAGAGGAC 1384
Db TGAATATCTTTATTTGACTTCCACATATTTCCAGTAATTTCTGTTATGAGAGGAC 223
1385 TTGAATAGCAAAATCCACACAGTACTGATAGATAGATAGATAGATAGATAGATAGAT 1444
Db TTGAATAGCAAAATTC-CACACAGTAACTGGATAGACACGATAGATAGATAGATAGAT 164
1445 CACTTGGTACTACACCGAACTCAAAATGTTCTTTCTCTGATGAGATAGATAGATAGAT 1504
Db CACTTGGTACTACACCGAACTCAAAATGTTCTTTCTCTGATGAGATAGATAGATAGAT 106
1505 CTTTGTGATAGTCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 1564
Db CTTTGTGATAGTCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCT 50
1565 TTCTATCCAGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1612
Db TTCTATCCAGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2
RESULT 13
AC111087/c 231508 bp DNA linear ROD 30-SEP-2003
LOCUS
DEFINITION
AC111087
VERSION
KEYWORDS
HTG
SOURCE
MUS musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1. (bases 1 to 231508)
REFERENCE
AUTHORS
TITLE
MUS musculus chromosome 9, clone RP23-152016, complete sequence.

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 231508)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lakocque, K., Lamaze, R.,
Landers, T., Lechoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 231508)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
McEwan, P., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (03-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 231508)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
McEwan, P., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission

source	1. 231508	"/organism="Mus musculus"	1. 231508	"/organism="Mus musculus"
		"/mol_type="genomic DNA"		"/mol_type="genomic DNA"
		"/db_xref="taxon:10090"		"/db_xref="taxon:10090"
		"/chromosome="9"		"/chromosome="9"
		"/map="9"		"/map="9"
		"/clone="RP23-152016"		"/clone="RP23-152016"
misc_feature	1. 20033	"/clone_lib="RPCI-23 Fema	1. 20033	"/clone_lib="RPCI-23 Fema
		"/note="was end extension		"/note="was end extension
unsure		"/clone_end_Sp6="		"/clone_end_Sp6="
		complement(1.. 96)		complement(1.. 96)
unsure		"/note="single clone cover		"/note="single clone cover
		complement(1573..1577)		complement(1573..1577)
unsure		"/note="<30 qual SNGL reg		"/note="<30 qual SNGL reg
		complement(1815..1849)		complement(1815..1849)
repeat_region		"/note="single clone cover		"/note="single clone cover
		complement(2101..2522)		complement(2101..2522)
repeat_region		"/rpt_family="MLTIG1"		"/rpt_family="MLTIG1"
		complement(2551..2714)		complement(2551..2714)
unsure		"/rpt_family="MLTIG1-int"		"/rpt_family="MLTIG1-int"
		2592..2596		2592..2596
unsure		"/note="<30 qual SNGL reg		"/note="<30 qual SNGL reg
		2738..2834		2738..2834
unsure		"/note="single clone cover		"/note="single clone cover
		3092..3096		3092..3096
repeat_region		"/note="<30 qual SNGL reg		"/note="<30 qual SNGL reg
		3100..3151		3100..3151
unsure		"/rpt_family="(TTTG)n"		"/rpt_family="(TTTG)n"
		3413..3418		3413..3418
unsure		"/note="<30 qual SNGL reg		"/note="<30 qual SNGL reg
		3432..3436		3432..3436
unsure		"/note="<30 qual SNGL reg		"/note="<30 qual SNGL reg

Qy	225	TAATTGAAAAGGGGATATAAAGCATTAATAATATATAGATATATAGCTATATCTATGTATGTA	284
Db	66685	TAACCTCATAGAAGAATGGAAATTTAAAGATATATGATATATACATGTATGTACACATATGTA	66626
Qy	285	TCTAACAGAGAAGTTTCAAGTCACTTCAATTTAAAGAAACATTTTTTTGAGCATGGGACCAGCC	344
Db	66625	TGTGTGTGTAGTTTCATATAATTTTCATAAATTTTAAATTTCTTAATTTCTTAAGCAACAGC	66566
Qy	345	TCAGGCTTTATGCTGGGATGCAGTAGACA - GGAGATGGGGAATTTAGAAAAAGAGAAGACTGT	403
Db	66565	CTGAGATCCTTGTGAAATGCGGAGTCAGGGGGAAGAGAGACTTCAAAATGAGAAAAATAC	66506
Qy	404	GTAATTGAATGACCTGGGCTGCACCTTATAGGAACTTATAATTAATGATGATCTGATATA	463
Db	66505	ATCTTTAAGCTGACTTGCCCTGTAGAGCTTTGGGACTTGCAATTAATTAATTGAGCTGGATG	66446
Qy	464	AAACAT-----ACCAGGATAAAGATGTCAAATGAGTGTGACTCCCTTAAAGTAGATTA	515
Db	66445	AATGTGAATATGGATCTAATGATGAATATCAAAAACCTGTAATGCCTTCAAAGTAGATGA	66386
Qy	516	AAGTGTGCATCTTTTGTGTTTCTTAAAAATGATTTTACTGCTTGAAATTTACATTTTGAGTTG	575
Db	66385	AAATATATCTTTTCATGTTTCCCAAATAGCATGTCCTGCAATGAAGTCAATTTTTCGATTTG	66326
Qy	576	AAAGTTTAGAAACTAACATAGCATTAATTAATGAATATGCAATGCAAAATTAATTAATCTTTGA	635

```
Db 66325 TA-----CTTAACTGTAAATATGTTGTTAAATAGTAATACCATGGA 66285
QY 636 AAACGATGATAAATATATTCCTCCCTCTCTAGAAACAGTCAAAAGCCACATTCACAA 695
Db 66284 AACCATATCTCTGTAACATTTTACCT-GTTTGGATATGCGTATGATGATTCATTTGA 66226
QY 696 GTTCCAAATATAAGGAGGTAGCAAGTATAGGCGATGATATATTTCTGCGTGTGTGT 755
Db 66225 GTTACAAA-----CAAGTCAGGCGAGGATAGATCTGTCCT----- 66189
QY 756 ATACCCATGCGCAGGCGCTTTTATAGGACTCCCAAAAGCATTTTGAAGATGGCAATAT 815
Db 66188 -----TCAAGGCCATCAACAAGCATTTTGGGTGAT-GAAGTAT 66153
QY 816 CAAATAGCTATGCTCTCTCAAAAGAGCGCATTTTAAATGTTACATATTTGCGAGC 875
Db 66152 TAAATCAGTGCATGTTTCTTCAACAG-GAAGGTTTAAATAGTAAATCCCAAGTACATAT 66094
QY 976 TCAGGTTATGATGTTTATGAATAAAGCTTCATTTCTTATAGCTACATCTATT 935
Db 66093 TAAAGTCTGGTATGTTTATTAAGTAAAGTATTTTATTTT-TCATTGATATGCTCTATT 66035
QY 936 ATTCCTTTTAGAACAAGATAACATTAAGTTTAAATAGTGGCATCTTAGCAT-TTA 994
Db 66034 ATTCCTTTT-----ACAGAGATAACAGTATATTTAAATCATTCGCTGTTTAGCATATTT 65979
QY 995 TCAGGCTCTAATGAACCAATATTTGATCTCTCATAAATATTTTCTGATGTTACTAGCTAT 1054
Db 65978 TCAGATCTAAGAACCATATTAATCTTGACAAATATTTTCATGGAGTCTGCGAGTCT 65919
QY 1055 GGAATTAAGACTGCGACCAACCTGACATTAAGTGGAAATGTTAGGATTTTTCGGC 1114
Db 65918 GGGAAACCAACATGCTGGAAGCTTGACTTAACCAAG-AGTTTATTAATTCGCGAGC 65860
QY 1115 ATCGCATGTTAGAT---CTCTAAATTTAAACATTCCTGTTAAATGCTAAGTTTCTCT 1171
Db 65859 ACATATGTTGAGTTCATATTAATATGACAAATTCATATTAAGTCTCATTTATGAC 65800
QY 1172 TTTATCAATGATTAATCTGAAGGCCAATATCATCAATTAATCAATGAAAGC 1222
Db 65799 TCTTCTTATCATTTTCTTAAAGCCAACTACATACATAAACAACAATAAC 65749

RESULT 14
AC116714
LOCUS
DEFINITION Mus musculus clone RP23-122E19, WORKING DRAFT SEQUENCE, 35
unordered pieces.
ACCESSION AC116714
VERSION AC116714.2 GI:30018030
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 256138)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-122E19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 256138)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,I., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramadamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuppback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 256138)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,P., Coram,B., DeArrellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramadamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuppback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi:19881957.
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22412
Center clone name: 122_E_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 57828: contig of 57828 bp in length
* 57829: gap of 100 bp
* 57929: contig of 624 bp in length
* 58552: gap of 100 bp
* 58553: contig of 27 bp in length
* 58653: gap of 100 bp
* 58679: gap of 100 bp
* 58779: contig of 620 bp in length
* 59399: gap of 100 bp
* 59499: contig of 100 bp
* 60497: contig of 998 bp in length
* 60597: gap of 100 bp
* 60598: contig of 677 bp in length
* 61274: gap of 100 bp
* 61375: contig of 977 bp in length
* 62351: gap of 100 bp
* 62451: contig of 1038 bp in length
* 63489: gap of 100 bp
* 63589: gap of 100 bp
```

TITLE
JOURNAL
REFERENCE
AUTHORSTITLE
JOURNAL
COMMENT


```
Db 17346 TAAATCAGTCGATGTTTCTTACCAAG-GAAAAGTTTAAATAGTTAAATCCCAAGTACATAT 17404
QY 876 TCAGGTTATGATGTTTATGAAAATAAGCTTCATTTTCTTATAGTACATACCTTATT 935
Db 17405 TAAAAGTCGTGATGTTTATGAAAATAAGCTTCATTTTCTTATAGTACATACCTTATT 17463
QY 936 ATTCCCTTTTGAACAACAAGATAACAATAAGTTTAAATAGTTGCAATCTAGCAT-TTA 994
Db 17464 ATTCCGTTT-----ACAGAGAATAACAGTATATTTAAATCAATTCGTGTATTTAGCATATTT 17519
QY 995 TCAGGCTTAATGAACCAACCAATATGAACTCTGATATAATTTTCTGATGTTTACTAGCTAT 1054
Db 17520 TCAGATCTTAAGAAACCAATTAATAATCTGTCGAAATATTTTCATGAGTCTCGAGTCT 17579
QY 1055 GGGAAATTAGAACTGGCAACACCCCTGACATTAAGTGGAAATGTTTAGGATTTTTCGGC 1114
Db 17580 GGGAAACCAACCAATGCTGGAAGCTTGACTTAACCAACAG-AGTTTATAATTTCTGGAGC 17638
QY 1115 ATCGCATGTTAGAAAT---CTCTAAATTTAAACATTCCTGTTAAATGACTAAGTTTGCT 1171
Db 17639 ACCATATGTTTGAGTTCATATTAATGATGAACAAATTCATATTAATGACTTCATATGAC 17698
QY 1172 TTATCAATATGAATTCGAAGCCCAATATCATACATTAACCTATGAAGC 1222
Db 17699 TCTTCTTATCATTTTCTTAAGGCCCAACACTACATATAACAACAATAAC 17749

RESULT 15
AC132694/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Muzny,D,Marie, Metker,M, Lee, Abranzon,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Georgievski,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunaratne,P, Haaland,W, Hamill,C, Hamilton,N, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpach,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenshew,L, Loulseged,H, Lozado,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindartne,M, Mahmoud,M, Malloy,K, Mangun,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwaokeme,O, Okwuonu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
```

Flopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shwartzbeyn, A., Sisson, J., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, F., Taylor, C., Taylor, R., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Walldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, B., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 202314)
Rat Genome Sequencing Consortium.
Submitted (03-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 202314)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23908401.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCCU
Center clone name: CH230-490C22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 169128 bases at least Q40
Consensus quality: 170817 bases at least Q30
Consensus quality: 171921 bases at least Q20
Estimated insert size: 173670; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved
* 1 202314: contig of 202314 bp in length.
Location/Qualifiers
1 .202314

FEATURES
source

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:00:48 ; Search time 1262.83 Seconds
(without alignments)

16458.288 Million cell updates/sec

Title: US-10-051-835-14

Perfect score: 696

Sequence: 1 taaatatctgatggcagt.....ttttgaaaatatataaaat 696

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rcd:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gsl1:*
 - 29: gb_gsl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	262.4	37.7	286	10	AW838776	CM4-LT005
C	2	67.6	9.7	997	29	CNS005TE	AL060767 Drosophil
C	3	61.2	8.8	987	29	CNS00418	AL066537 Drosophil
C	4	55.4	8.0	1101	29	CNS0039G	AL063921 Drosophil

C	5	54.8	7.9	846	28	B18548	B18548 T408-Sp6 TA
C	6	50.8	7.3	1373	29	CNS01637	AL106237 Drosophil
C	7	49.8	7.2	1101	29	CNS000D1	AL065414 Drosophil
C	8	49.4	7.1	1101	29	CNS017RP	AL108415 Drosophil
C	9	48.6	7.0	1101	29	CNS0181N	AL108773 Drosophil
C	10	48.2	6.9	687	29	CE833588	CE833588 tigr-g88-
C	11	48.2	6.9	878	29	CNS0187R	AL108993 Drosophil
C	12	47.8	6.9	1478	12	BM414951	BM414951 OP20019 M
C	13	47.6	6.8	932	9	AL514901	AL514901
C	14	47.6	6.8	994	13	BM414650	BM414650
C	15	47.4	6.8	1201	13	BM414650	BM414650
C	16	46.8	6.7	1101	29	CNS0039L	AL063926 Drosophil
C	17	46.6	6.7	986	14	CD387197	CD387197 AGENCOURT
C	18	46.6	6.7	1092	13	BM359398	BM359398
C	19	46.6	6.7	1201	13	BM360615	BM360615
C	20	46.6	6.7	1201	13	BM387747	BM387747
C	21	46.4	6.7	1101	29	CNS0039R	AL063932 Drosophil
C	22	46.4	6.7	1201	13	BM375239	BM375239
C	23	46.2	6.6	617	28	AZ077015	AZ077015 RPCI-23-4
C	24	46.2	6.6	854	12	BM414967	BM414967
C	25	46	6.6	1101	29	CNS00F86	AL070260 Drosophil
C	26	45.8	6.6	976	13	BM411173	BM411173
C	27	45.8	6.6	1109	14	CA453621	CA453621 AGENCOURT
C	28	45.6	6.6	970	13	BM349797	BM349797
C	29	45.6	6.6	1101	29	CNS0182P	AL108811 Drosophil
C	30	45.4	6.5	865	12	BG343667	BG343667 HVSMEG000
C	31	45.4	6.5	1201	13	BM361080	BM361080
C	32	45.4	6.5	1209	29	CG749482	CG749482 F043-3-H1
C	33	45.2	6.5	1201	13	BM385572	BM385572
C	34	44.8	6.4	982	29	CNS016V1	AL107239 Drosophil
C	35	44.8	6.4	1035	29	CNS04JPG	AL293821 Tetraodon
C	36	44.8	6.4	1101	13	BM439822	BM439822
C	37	44.8	6.4	1101	29	CNS0006J	AL062049 Drosophil
C	38	44.8	6.4	1101	29	CNS017KX	AL108171 Drosophil
C	39	44.8	6.4	1201	13	BM355642	BM355642
C	40	44.6	6.4	313	9	AA370340	AA370340 EST82022
C	41	44.6	6.4	954	13	BQ670925	BQ670925 AGENCOURT
C	42	44.6	6.4	954	13	BM326810	BM326810
C	43	44.6	6.4	1028	13	BM431780	BM431780
C	44	44.6	6.4	1201	9	AL538931	AL538931
C	45	44.6	6.4	1201	13	BM361533	BM361533

ALIGNMENTS

RESULT 1	AW838776/c	AW838776	286 bp	mRNA	linear	EST 18-MAY-2000
LOCUS	CM4-LT0057-140100-080-h05 LT0057	Homo sapiens	cdna, mRNA	sequence.		
DEFINITION	AW838776	AW838776				
ACCESSION	AW838776.1	GI:79327250				
VERSION	EST.					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 286)					
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,W.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.					
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)		
MEDLINE	20202663					
PUBMED	10737800					
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,					

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=4t2=CM4-LT0057-140
100-080-h05kt3=2000-01-14&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 4
High quality sequence stop: 286.

FEATURES

Location/Qualifiers
1..286
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="LT0057"
/note="Organ: leiomyos; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 37.7%; Score 262.4; DB 10; Length 286;
Best Local Similarity 99.6%; Pred. NO. 5.1e-42;
Matches 263; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 396 ATGAAGAAGGTGACACAGCCATTCTCTTACACAGAAATACAGATTTCACAGCT 455
Db 286 ATGAAGAAGGTGACACAGCCATTCTCTTACACAGAAATACAGATTTCACAGCT 227
QY 456 CTAAACAAAGAAACAAAGTTGGCAACATCTCTCTTCTGAATATGAAGAGAGGG 515
Db 226 CTAAACAAAGAAACAAAGTTGGCAACATCTCTCTCTGAATATGAAGAGAGGG 167
QY 516 GATAAATCTGGAGTAGGATTGAAAAAGTCAAGAGAAAAAAGAACAGCCCAAGT 575
Db 166 GATAAATCTGGAGTAGGATTGAAAAAGTCAAGAGAAAAAAGAACAGCCCAAGT 107
QY 576 GTAAACAGATATCTCTCCATGGGATGTAAGAAAGGAGTATTTCATATAAATGACCCCTT 635
Db 106 GTAAACAGATATCTCTCCATGGGATGTAAGAAAGGAGTATTTCATATAAATGACCCCTT 47
QY 636 GGAAGGAGTTCAACAGAGTTGCC 659
Db 46 GGAAGGAGTTCAACAGAGTTGCC 23

RESULT 2

CNS005TE/c
LOCUS 997 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR1222 of RCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
REFERENCE
AUTHORS Direct Submission
TITLE Genoscope - Centre National de Sequencage :
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
EP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Oosawa and
Aaron Mammeter in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1..997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR1222"
/clone_lib="RCI-98"
/note="end : TET3"

ORIGIN

Query Match 9.7%; Score 67.6; DB 29; Length 997;
Best Local Similarity 17.2%; Pred. No. 0.0014;
Matches 113; Conservative 198; Mismatches 346; Indels 0; Gaps 0;
QY 36 GGAACACAGAGAGAGGCTTTGATGGGATACAGACTTGGAGACATCATGCTGAGCAG 95
Db 992 RRGARARARRRRRRRRRRGARARARRRRRRRRGARARARRRRRRRRGARARAR 933
QY 96 TAAATGAGATGATTCAGAAAGAGTATAAATCTGGGAGAGAGACAGAGCCTCAAGG 155
Db 932 RAARARARRRRRRRRRRRRGARARRRRRRRRRRGARARARRRRRRRRRRR 873
QY 156 AACATATTTAAGACTGGGTAGAAAACAAAGAGATGATGAACAAGTCTAGGAGATTAT 215
Db 872 RAKAARARGARRRRRRRRRRRRRGARAGGAGRRRRRRRRRRRRRRRRRR 813
QY 216 AGCAGTCACTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGAGCCAGATTTCAGCAT 275
Db 812 RGRARARARRRRRRRRRRRRGARRRRRRRRRRRRRRRRRRRRRRRRR 753
QY 276 CAGAACCATCAAGCATTTTGGGGTGAAGAAAGAGCCATGATCAAAAGTGGGGA 335
Db 752 AGAFAAARRRRRRAGAAARRRRRRRRRRRRRRRGAGARARRRRRRRRRRR 693
QY 336 AAAGGCCTTTTGGAGTGGCAGTGTGGGTAGGGAGTAGGGCTCCGTTAGAATACATGG 395
Db 692 GRRGRGARRRRRRRRRRRGAGARRRRRRRRRRRRRRRRRRRRRRRRRR 633
QY 396 ATGAAGAAGGTGAGCACAGCCATTTCCTTTACACAGAAATAACAGATTTCAGCCT 455
Db 632 RRR 573
QY 456 CTAAACAAAGAAACAAAGTTTGGGAACATCTCTCTCTCTGAATATCAAGAGAGGG 515
Db 572 RRGARARARRR 513
QY 516 GATAAATCTGGAGTAGGATTGTGAAAAAGTCAAGAGAAAAAAGAACAGCCCAAGT 575
Db 512 RRGARR 453
QY 576 GTAACAGATCTTCTCCATGGGTGTAAGAAAGGAGTATTTCATATAAATGACCCCTT 635
Db 452 NNNNNNNAGCCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 393
QY 636 GGAAGGAGTTCAAAACAGGTTGCCATATGCTTATGTGGAGTTTGGAAAAATATATA 692
Db 392 NNN 336

```

RESULT 3
CNS00418/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL066537
VERSION
  AL066537.1 GI:4942778
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 987)
REFERENCE
  Direct Submission
  AUTHORS
  TITLE
  JOURNAL
  COMMENT
    Determination of this BAC-end sequence was carried out as part of a
    collaboration with the Berkeley Drosophila Genome Project (BDGP).
    The BDGP is constructing a physical map of the Drosophila
    melanogaster genome using these BACs. For further information
    please see http://www.fruitfly.org The BDGP Drosophila
    melanogaster BAC library was prepared by Kazutoyo Osoegawa and
    Aaron Mammosser in Pieter de Jong's laboratory in the Department of
    Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
    NY. The library is named RPCI-98 and was constructed by partial
    EcoRI digestion of Drosophila DNA provided by the BDGP from the
    isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
    P1 and EST libraries. A more detailed description of the library
    and how to order individual BAC clones, the entire library, or
    filters for hybridization from the BACPAC Resource Center can be
    found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
  Location/Qualifiers
  1..987
  /organism="Drosophila melanogaster"
  /mol_type="genomic DNA"
  /db_xref="taxon:7227"
  /clone="BACR09C16"
  /clone_lib="RPCI-98"
  /note="end : TET3"

ORIGIN
  Query Match      8.8%; Score 61.2; DB 29; Length 987;
  Best Local Similarity 22.0%; Pred. No. 0.027;
  Matches 90; Conservative 136; Mismatches 184; Indels 0; Gaps 0;

  QY 2 AAATATCTGATGACCTAGCAATTTGAGTTTGGAAACACAGAGAGAGCTTTGATGG 61
  Db 917 AARAPAAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAAR 858
  QY 62 CGATACAGACTTGGAGACATCAGTCTGACAGCTAAATGATGATTCAGGAAGAGTA 121
  Db 857 RRRRGRGRGAGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 798
  QY 122 TAAATCTGGAAGACACAGAGCAGGCTCAAGGAAATATTTAAGGACTGGGTAGAAAA 181
  Db 797 AEGRRGRGARGRGGAARAAARAAARAAARAAARAAARAAARAAARAAARAAAR 738
  QY 182 ACAAGAGATATGAACAGAGTGGAGATTTATGACAGTGCACCTTTGAGAGTACATCTC 241
  Db 737 AAARAAAGAAAAARAAARAAARAAARAAARAAARAAARAAARAAARAAARAAAR 678
  QY 242 TAGAGTGTATGTGAGAGCAGAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGTG 301
  Db 677 GGGRRRGAGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 618
  QY 302 GAAGAAAAGAGCCATGAATCAAAAGTGGGGAAGAGCCCTTTTGGGAGGTGCGAGTGT 361
  Db 617 ARRGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 558

```

```

362 GGCTAGGAGTAGGCTCGGTTAGATACATGATGATGAAGAAAGTGTGAG 411
Db 557 GGGGGRGRRGGGGGGGGGRRGRRGRRGRRGRRGRRGRRGRRARAR 508

RESULT 4
CNS0039G/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL063921
VERSION
  AL063921.1 GI:4941778
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)
REFERENCE
  Direct Submission
  AUTHORS
  TITLE
  JOURNAL
  COMMENT
    Determination of this BAC-end sequence was carried out as part of a
    collaboration with the Berkeley Drosophila Genome Project (BDGP).
    The BDGP is constructing a physical map of the Drosophila
    melanogaster genome using these BACs. For further information
    please see http://www.fruitfly.org The BDGP Drosophila
    melanogaster BAC library was prepared by Kazutoyo Osoegawa and
    Aaron Mammosser in Pieter de Jong's laboratory in the Department of
    Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
    NY. The library is named RPCI-98 and was constructed by partial
    EcoRI digestion of Drosophila DNA provided by the BDGP from the
    isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
    P1 and EST libraries. A more detailed description of the library
    and how to order individual BAC clones, the entire library, or
    filters for hybridization from the BACPAC Resource Center can be
    found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
  Location/Qualifiers
  1..1101
  /organism="Drosophila melanogaster"
  /mol_type="genomic DNA"
  /db_xref="taxon:7227"
  /clone="BACR08K10"
  /clone_lib="RPCI-98"
  /note="end : TET3"

ORIGIN
  Query Match      8.0%; Score 55.4; DB 29; Length 1101;
  Best Local Similarity 16.2%; Pred. No. 0.37;
  Matches 110; Conservative 284; Mismatches 284; Indels 1; Gaps 1;

  QY 13 TAGGCAGTTAGAAATTTGAGTTTGGAAACACAGAGAGAGCTTTGATGGCATACAGACT 72
  Db 1100 KARRWGDDTTDRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTRDTR 1041
  QY 73 TGAAGACATCAGTCTGACAGCTAAATGAGATGATTCAGGAAGAGATATAAAGTGGAA 132
  Db 1040 WWWATWTWTDKWWWWATAAKTDTATWTWTWTATWADWAGDRGAGKEDRDATDAGA 981
  QY 133 GAGGACAGAGGACAGCTCAAGGACATATTTAAGCATCGGTGAGAAACAAAGAGTA 192
  Db 980 GRDGGRRKDKKRRKGGDDDKKGGKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 921
  QY 193 TGAACAGAGTGGAGAGATTTATTAGCAGTGACCTTTGAGAGTACATCTTACAGTGGTAT 252
  Db 920 DDDDDG- GNDDDGKDDKDDDDTGTGDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDD 862
  QY 253 GTGTAGAGCCAGATTTTACATCAGAACCATCAAGCATTTTGGGGGTGGAAGGAAGG 312
  Db 861 GWADADWTTWDAADWDWADWADWADWADWADWADWADWADWADWADWADWADWADWAD 802

```


ORIGIN	Query Match	6.9%	Score 47.8;	DB 12;	Length 1478;
	Best Local Similarity	40.4%;	Pred. No. 11;		
	Matches 133;	Conservative 0;	Mismatches 196;	Indels 0;	Gaps 0;
Qy	116	AGAGTAAACTGGGAAGAGCAGACAGAGGACAGGCTCAAGGAAACATATTTTAAGACATGGGT	175		
Db	760	ANAGAGAAANANAGAGGGGACAGGAGAGAGAGAAAAAAGAAAAAGGAGGNGGGG	701		
Qy	176	AGAAACACAGACAGATATGAACAGAGTGAGGAGATATTATTAGCAGGTGACCTTTTGAGAGTA	235		
Db	700	AANNAAAGACANGANGNAGGAGAGGAGAAANAAAAAAGGNGNGCGNNNNNNNN	641		
Qy	236	CATCTCTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTG	295		
Db	640	CCNNNNAAAAAANNNNNNTTTTTTGAGGGGATTTTTTTTTTTTAAAAAGNGAGAGAG	581		
Qy	296	GGGGTGGAAAGGAAAAAGGAGCCATGAATCAAAAGGTGGGAAAAAGGCCCTTTTGGGAGGTGG	355		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:43:48 ; Search time 203.468 seconds
(without alignments)
15422.560 Million cell updates/sec

Title: US-10-051-835-14
Perfect score: 696
Sequence: 1 taaataatctgtaggcagct.....ttttgaaaaatatataaat 696

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 225433464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	696	15	US-10-051-835-14
2	46.4	6.7	461	10	US-09-814-353-17724
3	44.8	6.4	9415	15	US-10-311-455-268
4	41.8	6.0	300	9	US-09-864-761-26934
5	41	5.9	508	10	US-09-814-353-18511
6	40.6	5.8	6668	15	US-10-311-455-1669
7	40.2	5.8	671	15	US-10-184-644-346
8	40.2	5.8	671	15	US-10-184-634-346
9	40.2	5.8	1269	13	US-10-282-122A-16543
10	40	5.7	767	13	US-10-027-632-153507
11	40	5.7	767	16	US-10-027-632-153507
12	39.6	5.7	6319	15	US-10-311-455-1616
13	39.4	5.7	2750	13	US-09-854-867-362
14	39.4	5.7	6681	15	US-10-311-455-128

C 15	39.4	5.7	6681	15	US-10-240-452-4	Sequence 4, Appli
C 16	39	5.6	583	13	US-10-027-632-136545	Sequence 136545,
C 17	39	5.6	583	16	US-10-027-632-136545	Sequence 136545,
C 18	38.8	5.6	921	13	US-10-027-632-120814	Sequence 120814,
C 19	38.8	5.6	921	16	US-10-027-632-120814	Sequence 120814,
C 20	38.8	5.6	10048	13	US-10-221-613-212	Sequence 212, Appli
C 21	38.8	5.6	3673778	15	US-10-312-841-1	Sequence 1, Appli
C 22	38.4	5.5	421	15	US-10-060-036-3868	Sequence 3868, Ap
C 23	38.4	5.5	657	13	US-10-276-774-703	Sequence 703, App
C 24	38.4	5.5	699	10	US-09-764-891-185	Sequence 185, App
C 25	38.4	5.5	718	13	US-10-027-632-12438	Sequence 12438, A
C 26	38.4	5.5	718	16	US-10-027-632-12438	Sequence 12438, A
C 27	38.4	5.5	1074	10	US-09-814-353-20743	Sequence 20743, A
C 28	38.4	5.5	1686	13	US-10-302-172-327	Sequence 327, App
C 29	38.4	5.5	2054	15	US-10-198-846-11116	Sequence 11116, A
C 30	38.4	5.5	14861	13	US-10-221-613-161	Sequence 161, App
C 31	38.4	5.5	14861	15	US-10-311-455-1167	Sequence 1167, Ap
C 32	38.2	5.5	610	13	US-10-027-632-261595	Sequence 261595,
C 33	38.2	5.5	610	16	US-10-027-632-261595	Sequence 261595,
C 34	38	5.5	5823	15	US-10-240-453-256	Sequence 256, App
C 35	37.8	5.4	5660	15	US-10-311-455-2330	Sequence 2330, Ap
C 36	37.6	5.4	6136	17	US-10-257-166-124	Sequence 124, App
C 37	37.6	5.4	9539	15	US-10-239-676-51	Sequence 51, Appl
C 38	37.6	5.4	9539	15	US-10-240-453-53	Sequence 53, Appl
C 39	37.6	5.4	14798	15	US-10-311-455-1006	Sequence 1006, Ap
C 40	37.6	5.4	96593	12	US-10-052-482-67	Sequence 67, Appl
C 41	37.4	5.4	410	15	US-10-060-036-3601	Sequence 3601, Ap
C 42	37.4	5.4	505	10	US-09-918-995-20057	Sequence 20057, A
C 43	37.4	5.4	2000	16	US-10-260-238-2483	Sequence 2483, Ap
C 44	37.4	5.4	7058	17	US-10-257-166-134	Sequence 134, App
C 45	37.4	5.4	370469	13	US-10-087-192-250	Sequence 250, App

ALIGNMENTS

RESULT 1
US-10-051-835-14
; Sequence 14, Application US/10051835
; Publication No. US20030165864A1
; GENERAL INFORMATION:
; APPLICANT: Jones, David A.
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES REGULATED BY DNA METHYLATION IN TUMOR CELLS
; FILE REFERENCE: PA-0044 US
; CURRENT APPLICATION NUMBER: US/10/051,835
; CURRENT FILING DATE: 2002-01-16
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165864A1 213413.1
US-10-051-835-14

Query Match	100.0%	Score	696	DB	15	Length	696
Best Local Similarity	100.0%	Pred. No.	7.3e-187	Mismatches	0	Indels	0
Matches	696	Conservative	0	0	0	Gaps	0
Qy	1	TAATAATCTGATAGGCAGTTAGAAATTTGAGTTTGGAAACACAGAGAGAGAGGCTTTGATG	60				
Db	1	TAATAATCTGATAGGCAGTTAGAAATTTGAGTTTGGAAACACAGAGAGAGGCTTTGATG	60				
Qy	61	GGATACAGACTTGGAGACATCATGCTGTCGACAGTAAATGAGATGATTCAGAGAAAGACT	120				
Db	61	GGATACAGACTTGGAGACATCATGCTGTCGACAGTAAATGAGATGATTCAGAGAAAGACT	120				
Qy	121	ATAAATCTGGAGAGAGACAGAGAGAGGCTCAAGGAACATATTTAAGGACTCGGTAGAAA	180				
Db	121	ATAAATCTGGAGAGAGACAGAGAGAGGCTCAAGGAACATATTTAAGGACTCGGTAGAAA	180				

```
QY 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
DB 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
QY 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
DB 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
QY 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
DB 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
QY 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
DB 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
QY 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
DB 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
QY 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
DB 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
QY 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
DB 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
QY 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
DB 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
QY 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696
DB 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696

RESULT 2
US-09-814-353-17724
; Sequence 17724, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Little, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-06-15
; PRIOR FILING DATE: 2000-07-07
; PRIOR FILING DATE: 2000-07-07
; PRIOR FILING DATE: 2000-07-25
; PRIOR FILING DATE: 2000-07-25
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 17724
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 195, 243, 244, 245, 397, 398, 399

QY 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
DB 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
QY 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
DB 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
QY 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
DB 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
QY 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
DB 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
QY 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
DB 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
QY 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
DB 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
QY 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
DB 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
QY 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
DB 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
QY 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696
DB 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696

OTHER INFORMATION: n = A, T, C or G
US-09-814-353-17724
Query Match 6.7%; Score 46.4; DB 10; Length 461;
Best Local Similarity 46.0%; Pred. No. 0.011;
Matches 143; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
QY 385 AGAATACATGGATGAAGAAGGTGAGCAGCAGCATTTCCCTTTTACACAGAAATAACAGA 444
DB 131 ATAATAAAAAATTTTAAAAAAATTTAAAAAAATTTAAAAAAATTTAAAAAAATTTT 190
QY 445 TTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGGGAACATTCCTCTCTCTGAAATAT 504
DB 191 TTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 250
QY 505 GAAAGAGAGGGGATTAATCTCTGGAGTAGGATTTGTGAAAAAGTCAAGAGAAAAAAGA 564
DB 251 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTT 310
QY 565 ACAGCCCAAGTGTAAACAGATCTTCTCCATGGGATGTTAAAAAGGAAGTTTATTTCA 624
DB 311 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTT 370
QY 625 AATGACCCCTTGAAGAGGATTTCAAAACAGGTTGCGCATATGCTTTATGTGAGTTTGA 684
DB 371 TTTTAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 430
QY 685 AATATATAAAAA 695
DB 431 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 441

RESULT 3
US-10-311-455-268/c
; Sequence 268, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 268
; LENGTH: 9415
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-268
Query Match 6.4%; Score 44.8; DB 15; Length 9415;
Best Local Similarity 49.2%; Pred. No. 0.16;
Matches 118; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
QY 387 AATACATGGATGAAGAAGGTGAGCAGCAGCATTTCCCTTTTACACAGAAATAACAGATT 446
DB 7617 AATAACCGTAAAAAATATATAATCGTAAAAAATTTCCCACTAATAAAAAA 7558
QY 447 TCCAGCCTCTAACCAAGAAAACAACAAGTTTGGGAACATTCCTCTCTCTGAAATATGA 506
DB 7557 ACTCCGCTCCAAACTAAAAAATTTAAAAAATTTAAAAAATTTTAAAAAATTTTAA 7498
QY 507 AAGAGAGGGGATTAATCTGGAGTAGGATTTGTGAAAAAGTCAAGAGAAAAAAGAAC 566
```

```
QY 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
DB 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
QY 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
DB 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
QY 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
DB 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
QY 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
DB 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
QY 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
DB 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
QY 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
DB 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
QY 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
DB 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
QY 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
DB 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
QY 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696
DB 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696

RESULT 2
US-09-814-353-17724
; Sequence 17724, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Little, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-06-15
; PRIOR FILING DATE: 2000-07-07
; PRIOR FILING DATE: 2000-07-07
; PRIOR FILING DATE: 2000-07-25
; PRIOR FILING DATE: 2000-07-25
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 17724
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 195, 243, 244, 245, 397, 398, 399

QY 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
DB 181 AACAGAGAGATGACAAAGAGTGGAGAGATTATTAGCAGTGACCTTTGAGAGTACATCT 240
QY 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
DB 241 CTAGAGTGGTATGTGTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGT 300
QY 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
DB 301 GGAAGAAAAGAGCCATGAATCAAAAGTGGGAAAAGCCCTTTTGGAGGTGGCAGTG 360
QY 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
DB 361 TGGTAGGAGTGGGCTCCGTTAGATACATGATGAAAGAGTGGGACACAGCCAT 420
QY 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
DB 421 TTCTTTTACACAGAAATAACAGATTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGG 480
QY 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
DB 481 GACATTCCTCTCTTCTGAATATGAAGAGAGGGGATTAATCTGGAGTAGATTGCA 540
QY 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
DB 541 AAAAAGTCAAGAGAAAAAAGAAAGACAGCCCAAGTGTAAACAGATCTTCCATGGGATG 600
QY 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
DB 601 GTAAAAGGAGTATTTCATTAATGATGACCCCTTGGAGGAGTTCAAAACAGTTGCCA 660
QY 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696
DB 661 TATGCTTTATGTGGAGTTTGGAAAAATATATAAAAT 696

OTHER INFORMATION: n = A, T, C or G
US-09-814-353-17724
Query Match 6.7%; Score 46.4; DB 10; Length 461;
Best Local Similarity 46.0%; Pred. No. 0.011;
Matches 143; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
QY 385 AGAATACATGGATGAAGAAGGTGAGCAGCAGCATTTCCCTTTTACACAGAAATAACAGA 444
DB 131 ATAATAAAAAATTTTAAAAAAATTTAAAAAAATTTAAAAAAATTTAAAAAAATTTT 190
QY 445 TTTCCAGCCTCTAACCAAGAAAACAACAAGTTTGGGAACATTCCTCTCTCTGAAATAT 504
DB 191 TTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 250
QY 505 GAAAGAGAGGGGATTAATCTCTGGAGTAGGATTTGTGAAAAAGTCAAGAGAAAAAAGA 564
DB 251 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTT 310
QY 565 ACAGCCCAAGTGTAAACAGATCTTCTCCATGGGATGTTAAAAAGGAAGTTTATTTCA 624
DB 311 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTT 370
QY 625 AATGACCCCTTGAAGAGGATTTCAAAACAGGTTGCGCATATGCTTTATGTGAGTTTGA 684
DB 371 TTTTAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 430
QY 685 AATATATAAAAA 695
DB 431 AAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAATTTTAAAAAAAT 441

RESULT 3
US-10-311-455-268/c
; Sequence 268, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 268
; LENGTH: 9415
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-268
Query Match 6.4%; Score 44.8; DB 15; Length 9415;
Best Local Similarity 49.2%; Pred. No. 0.16;
Matches 118; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
QY 387 AATACATGGATGAAGAAGGTGAGCAGCAGCATTTCCCTTTTACACAGAAATAACAGATT 446
DB 7617 AATAACCGTAAAAAATATATAATCGTAAAAAATTTCCCACTAATAAAAAA 7558
QY 447 TCCAGCCTCTAACCAAGAAAACAACAAGTTTGGGAACATTCCTCTCTCTGAAATATGA 506
DB 7557 ACTCCGCTCCAAACTAAAAAATTTAAAAAATTTAAAAAATTTTAAAAAATTTTAA 7498
QY 507 AAGAGAGGGGATTAATCTGGAGTAGGATTTGTGAAAAAGTCAAGAGAAAAAAGAAC 566
```


Db 7497 AAAAAACCGAAAAAATAAACCAAAAAATAAAAAACCGAAATTAATAAAAAA 7438
QY 567 AGCCAAAGTGTACAGACTCTTCCTATGGATGGTAAAAAGGAAGTATTTCAATAAAA 626
Db 7437 CGACCAAACTAACCACTTCTTAATACTAAAAATAAAAAAATACTAACTAAAA 7378

RESULT 4
US-09-864-761-26934/c
; Sequence 26934, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26934
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012079.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97

; OTHER INFORMATION: EST HUMAN HIT: AW589784.1, EVALUATE 1.10e+00
; OTHER INFORMATION: NT HIT: L11910.1, EVALUATE 4.00e-02
US-09-864-761-26934
Query Match 6.0%; Score 41.8; DB 9; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.17;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;
QY 3 AAATATCTGATAGCGAGTTAGAAATTTTCAGTTTGGAAACACAGAGAGAGGCTTTTCATGCG 62
Db 228 AAATGTGAAGTTTGCAGTTGAATATTTTGTGCTGGAATTTAGGAAAAACAATGTGTGGGC 169
QY 63 GA--TACAGACTTGAAGACATCAGTGTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
Db 168 CATCTGCATATCGGTGATGAATGAAGCCACAGAACCAATCAATCGAGCAAGAGT 109
QY 121 ATAACTGGAGAGACAGAGGAGGCTCAAGGAACATATTTTAAAGACTGGGTAGAAA 180
Db 108 AAATAAAGTGAAGAAAGGAGAGGAAGAAAGAGAGAAAGCAATGTTTAAAGG 49
QY 181 AACAAGAGAGTATGAACAAGAGTGAAGAGATTTATTAGCA 219
Db 48 AATGATAGTGATTTCAAATTTGCAAAAGAGAGACTGAGAA 10
RESULT 5
US-09-814-353-18511/c
; Sequence 18511, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-0063
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18511
; LENGTH: 508
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-18511
Query Match 5.9%; Score 41; DB 10; Length 508;
Best Local Similarity 52.7%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 399 AAAGAAAGGTGAGCAGCAGCCATTTCTTTACACAGAAATACAGATTTCCAGCCTCTA 458
Db 280 AAAAAAGGTTTACACAAATTTTTTTTCCAAAAAGTTTTTCCCCCCCCAAAA 221
QY 459 ACCAAAGAAACAAAGTTTGGGAACATCTCTCTTCTGAATATGAAGAGGGAT 518
Db 220 AAAAAAAGGTTTACACAAATTTTTTTTCCAAAAAGTTTTTCCCCCCCCAAAA 161
QY 519 AAATCTGAGTAGGATTTGTGAAAAAGTCAAGAGAAAAAAGCAACA 567


```

; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-346

Query Match      5.8%; Score 40.2; DB 15; Length 671;
Best Local Similarity 8.6%; Pred. No. 0.76;
Matches 46; Conservative 172; Mismatches 317; Indels 0; Gaps 0;

QY 34 TTGGAACACAGAGAGAGCGTTTGATGGCGATACAGACTTGGAGACATCATGCTGAGC 93
DB 70 KENKKRQNEFLWEIQNNPHASYSAPPVPPVSSDSSEAPENPADGSDADEDDERGVNAV 129
QY 94 AGTAAATGAGATGATTTCAGGAAGAGTATAACTGGGAAGAGGACAGAGGACAGCTCAA 153
DB 130 AVTATAASDRMESDSKSSDNSGLKKTPTALKMSVSKRKAKSSDLDOASVSPSEEN 189
QY 154 GGAACATATTTAAGACTGGGTAGAAAACAGAGAGATGAAACAGAGTGAGGAGATTA 213
DB 190 SSSSESEKTSQDFTPEKKAAPRRRPLGGRKKKAPASDSKSDSKADSDGAKPEPVA 249
QY 214 TTAGCAGTGACCTTTGAGAGTACATCTCTAGAGTGGTATGTGTAGAGCCAGATTTTCAGC 273
DB 250 MARSSSSSSSSSSSDSVSVKPPRGRKPAEKPLPFRGRKPKRPPSPSSSSSDSDE 309
QY 274 ATCAGAACCATCAAGCATTTTGGGGTGAAGGAAAGAGGACCATGAATCAAAAGGTGG 333
DB 310 VDRISWKRRDARRELEARRERREBEELRLREQEKEKERRERARRDRGEAERGSGS 369
QY 334 GAAAGGCGCTTTTGGAGGTGGCAGTGTGGTAGGAGTAGGCTCCGTTAGATACAT 393
DB 370 SGDELRDEDEPVKKRGRGRGPPSSSDSEPELEAREKKSAPKPOSSSTEPARKPGQ 429
QY 394 GGATGAAAGAAAGGTGAGCAGACAGCCATTTCTTTTACACAGAAATACAGATTTCCAGC 453
DB 430 KEKRVPEEKQAKPVKVERTKRSEFSDMRKVEKKEPSVEEKLQHLSEIKFALKVD 489
QY 454 CTCCTAACCAAGAAACACACAGTTTGGGACATCTCTCTCTCAATATGAAAGAGAG 513
DB 490 SPDVKRCLNLEELGTQLVTSOILQKNDIVATLKKIRRYKANKDVMKAEVYTRLSR 549
QY 514 GGGATAAATACATCGAGTAGGATGTGCAAAAAGTCAAGAGAAAAAAGAACAG 568
DB 550 VLGPXIEAVQKYNKAGMEKEKAEKLAGELAGEEAPQEKEDKFTDLSAPVNG 604

```

```

RESULT 9
US-10-282-122A-16543
; Sequence 16543, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carl, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16543
; LENGTH: 1269
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-282-122A-16543

Query Match      5.8%; Score 40.2; DB 13; Length 1269;
Best Local Similarity 47.8%; Pred. No. 1.1;
Matches 151; Conservative 0; Mismatches 163; Indels 3; Gaps 1;

QY 19 GTTGAATAATTGAGTTTGGAAACACAGAGAGAGGCTTTGATGGCGATACAGACTTGAAG 78
DB 400 GCITTAATATAGCTTATAACATTCAAGTGGAGACTTTTATAGCCCTATATGATGAGAT 459
QY 79 ACATCAGTCTGACGAGTAAATGAGATGATTACGGAAGAGTATAACTGGGAGAGGAC 138
DB 460 AATACTCCAGATAAAATGATTTAAATATCTTATGAAAACATAATTTGAGGACGAGGT 519
QY 139 AGAGGACAGCTCAAGGAACATATTTTAAAGACTGGGTAGAAAAACAGAGAGTATGAACA 198
DB 520 TTAGGACGAGTTATAGGAA--AGTTTGAACATAGAAATAAGACAGAAATATGTTAACA 576
QY 199 AGAGTGAAGGAGATTTAGCAGTGCCTTTGAGAGTACATCTCTAGAGTGGTATGCTAG 258
DB 577 AGATTTTATAATATAGAGACCTTTAAGTTTCAATGATGTGCCACAGAGAGATGGAAG 636
QY 259 AGCCAGATTTTACGATCAGAACCATCAAGCATTTTGGGGTGGAGGAAAGAGGCCAT 318
DB 637 TTATTAATTTTGTACTATACCTGCGACTAATTTTGTAGTAAGAAGATATATACAA 696
QY 319 GAATCAAAAGTGGGGA 335
DB 697 GAGTTGAATGGATGGGA 713

```

```

RESULT 10
US-10-027-632-153507/c
; Sequence 153507, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

```

PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 153507
LENGTH: 767
TYPE: DNA
ORGANISM: Human
US-10-027-632-153507

Query Match
Best Local Similarity 5.7%; Score 40; DB 13; Length 767;
Matches 133; Conservative 0; Mismatches 135; Indels 2; Gaps 1;

QY 381 GGTAGATACATGATGAAGAAGGTGAGCACAGCCATTTCTTTACACAGAAATAA 440
DB 651 GGCTATAATATGTTGATCAATCAGTTGTAATGCAATTCATTTTGA--GAAACAGC 594
QY 441 CAGATTTCCAGCCTCTAACCAAGAACCAACAAGTTTGGACATTCCTCTCTGAA 500
DB 593 CCTCTTTCTGGTGTGTTTCAAGAGATATGGAAGTACTGACCCATTTTAAACCTGG 534
QY 501 ATATGAAGAGAGGGGATAAATACTGGAGTAGGATTTGAAAAAGTCAAGAGAAAAA 560
DB 533 ACATAAAAGGACCGTATATGATGAATAATGTTAGGAAAAAATCACCTATAGAAAAG 474
QY 561 AAGACAGCCCAAGTGAACAGATCTTCCATGGGATGTAAGAGGAAGTTATTCA 620
DB 473 AATGGAATGTCAAGTTCTCTCTGTAATGGTATGATGCTTCAAGTGAATTTAGCCA 414
QY 621 ATAAATGACCCCTTGAAGGAGTTCAAAA 650
DB 413 ATAAAGTCCATGCCCTTAAGAAGTTTATAA 384

RESULT 11
US-10-027-632-153507/c
Sequence 153507, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 153507
LENGTH: 767
TYPE: DNA
ORGANISM: Human
US-10-027-632-153507

RESULT 12
US-10-311-455-1616/c
Sequence 1616, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEFENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1616
LENGTH: 6319
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1616

Query Match
Best Local Similarity 5.7%; Score 39.6; DB 15; Length 6319;
Matches 108; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 427 TACAACAGAAATAACAGATTTCCAGCGCTCTAACCAAGAAACACCAAGTTTGGAAACAT 486
DB 705 TAAAAACATAAAAAATAACAAACAAAAAATAAAAAAATAAAAAAATAAAAAA 646
QY 487 TCCTCTCTTCTGAAATATGAAGAGAGGGGATAAATACTGGAGTAGGATTTGAAAAAG 546
DB 645 ACAATCTCTTACGAAATATAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 586
QY 547 TCAAGAGAAAAAAGAAAGAACAGCCCAAGTGTAAACAGATCTTCTCCATGGGATGGTAAA 606
DB 585 AAAAATCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 526
QY 607 AGGAAGTTATTTCAATAAATGACCCCTTGAAGAGGTTCAA 648
DB 525 ACTAACTAAATTCCTAAAAAATCTTCCCTTATCATCTCATCAA 484

LOCATION: (2099)...(2099)
OTHER INFORMATION: n is a, c, g or t
US-09-854-867-362

Query Match 5.7%; Score 39.4; DB 13; Length 2750;
Best Local Similarity 67.0%; Pred. No. 2.7;
Matches 69; Conservative 1; Mismatches 32; Indels 1; Gaps 1;

QY 3 AAATATCTCATAGGCAAGTTAGAAAATTGAGTTTGGAAACACAGGAGAGAGG-CTTTGATCG 61
DB 1743 AAATGTCGAGTAGGCGAGTTGGATATAGAGTCTGGARTTCAGGGAGAGGTCCTGGGCTCG 1684

QY 62 CGATACAGACTTGGAAAGACATCAGTCTCGAGCAGTAATGAGA 104
DB 1683 AGATATAAAATTTGGGAGTCATCAGCGTATAGATGTTATTTAAA 1641

RESULT 14
US-10-311-455-128/c
Sequence 128, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT FILING DATE: 2002-12-16
CURRENT APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 128
LENGTH: 6681
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-128

Query Match 5.7%; Score 39.4; DB 15; Length 6681;
Best Local Similarity 46.3%; Pred. No. 4.4;
Matches 130; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 413 ACAGCCATTTCTTTACACAGAAATAACAGATTTCCTCCAGCCTCTAACCCAAAGAAACAC 472
DB 5867 ATATCTCTTTCTTTAAAAACAAAAACAAAAATAATTAATACTAAAAATAAAAAATATC 5808

QY 473 AAGTTTGGACATTCCTCTCTGAAATATGAAGAGAGGGGATAAATACTGGAGTAG 532
DB 5807 AATAAAAAATTTAAAAAAATATTAATAATTTAAAAAAATACATATAACATAAATACTAA 5748

QY 533 GATTGTGAAAAAGCTCAAGAGAAAAAAGAACAGCCCAAGTGTAAACAGATCTTCTCC 592
DB 5747 ABACTTAACCTATTACTAAAAATAACAAAAAATCTTAATAATAATACTAATCCCAATCT 5688

QY 593 ATGGATGTTAAAAAGGAAGTTATTTCAATAAATGACCCCTTGGAGGAGTTCAAAACA 652
DB 5687 ACAAATTAACATCTTAATAATTAATAATTAATAAATAAATAAATAAATAAATAAATAA 5628

QY 653 GGTGGCCATATGCTTTATGTGGAGTTTGTAAAAATATATAA 693
DB 5627 ACATCTTAACATATTTATATATATATATATATATATATTA 5587

RESULT 15
US-10-240-452-4/c
Sequence 4, Application US/10240452

US-09-854-867-362/c
Sequence 362, Application US/09854867
Publication No. US20030224356A1
GENERAL INFORMATION:
APPLICANT: JOAN, KNOLL H
APPLICANT: ROGAN, PETER K
TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
FILE REFERENCE: 30307
CURRENT APPLICATION NUMBER: US/09/854,867
CURRENT FILING DATE: 2003-05-08
NUMBER OF SEQ ID NOS: 613
SOFTWARE: PatentIn version 3.1
SEQ ID NO 362
LENGTH: 2750
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: repeat region
LOCATION: (1)...(2750)
OTHER INFORMATION: line2
FEATURE:
NAME/KEY: misc_feature
LOCATION: (17)...(17)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (89)...(89)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (116)...(116)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (433)...(433)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (500)...(500)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (548)...(548)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (578)...(578)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (604)...(604)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (709)...(709)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (772)...(772)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (797)...(797)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1117)...(1118)
OTHER INFORMATION: n is a, c, g or t
FEATURE:
NAME/KEY: misc_feature

```
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240,452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 4
; LENGTH: 6681
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-4

Query Match      5.7%; Score 39.4; DB 15; Length 6681;
Best Local Similarity 46.3%; Pred. No. 4.4;
Matches 130; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 413 ACAGCCATTTCCTTACACAGAAATACAGATTCCCGCCTCTAACCCAGAAACAAC 472
Db 5867 ATATCTCTTTCTTTAAAAAACAACAAAAATAATTAATAAAAAATAAAAAATAC 5808

Qy 473 AAGTTTGGGAACATCTCTCTCTGAAATATGAAGAGAGGGGATAAATACCTGGAGTAG 532
Db 5807 AATAAAAAAATTTAAAAAATAATTAATAATTTAAAAAATAATACATAAACTAA 5748

Qy 533 GATTGTGAAAAAGTCAAGAGAAAAAAGAACAGCCCAAGTGTAAACAGATACCTCTCC 592
Db 5747 AAAACTAAACCTATTACTAAAAAATAACAAAAAATCTTAAATAATACTAATCCCAATCT 5688

Qy 593 ATGGGATGGTAAAAAGGAAGTTATTTCAATAAAATGACCCCTTGGAGGAGTTCAAAACA 652
Db 5687 ACAAAATTACATCTAATAATTAATAATAATTAATAACAAAAATTAATAATATCTAATAATTA 5628

Qy 653 GGTTGCCATATGCTTTATGTGGAGTTTGTGAAAAATATATAA 693
Db 5627 ACATCTTAACATATTTATATATATATTTATATATAATTAATTA 5587
```

Search completed: April 26, 2004, 03:19:16
Job time : 207.468 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2004, 07:15:18 ; Search time 34.5277 Seconds
(without alignments)
11186.557 Million cell updates/sec

Title: US-10-051-835-14
Perfect score: 696
Sequence: 1 taaatatactgagggcagt.....ttttgaaaaatatataaat 696

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	89	12.8	7218	1	US-08-232-463-14
C 2	40.8	5.9	6305	3	US-08-687-580B-6
C 3	37.8	5.4	604	4	US-09-334-818A-8
C 4	37	5.3	289	3	US-09-007-005-17
C 5	37	5.3	289	3	US-09-244-796-17
C 6	36.4	5.2	2089	1	US-08-552-142A-1
C 7	36.4	5.2	2089	1	US-08-910-973-1
C 8	36.4	5.2	2089	4	US-09-499-227-1
C 9	36.4	5.2	2089	5	PCT-US95-05741-1
C 10	36.2	5.2	832	4	US-09-621-976-2813
C 11	36.2	5.2	1664976	4	US-08-916-421B-1
C 12	36	5.2	1821	4	US-08-477-831C-1
C 13	36	5.2	1885	4	US-08-477-831C-9
C 14	36	5.2	1896	4	US-08-477-831C-10
C 15	36	5.2	1961	4	US-08-477-831C-8
C 16	36	5.2	2968	4	US-08-477-831C-13
C 17	36	5.2	3044	4	US-08-477-831C-12
C 18	36	5.2	392000	4	US-10-027-983-11
C 19	35.8	5.1	3084	4	US-09-620-312D-616
C 20	35.4	5.1	505	4	US-09-621-976-15639
C 21	35.4	5.1	6124	4	US-08-213-419B-3
C 22	35	5.0	248	3	US-09-007-005-32
C 23	35	5.0	248	3	US-09-244-796-32
C 24	35	5.0	277	3	US-09-007-005-3
C 25	35	5.0	277	3	US-09-244-796-3
C 26	35	5.0	1830121	4	US-09-557-884-1
C 27	35	5.0	1830121	4	US-09-643-990A-1

C 28	34.8	5.0	1166	4	US-09-072-596-323	Sequence 323, Appl
C 29	34.8	5.0	1166	4	US-09-072-967-328	Sequence 328, Appl
C 30	34.8	5.0	2502	2	US-09-234-332-1	Sequence 1, Appl
C 31	34	4.9	2853	3	US-08-589-711-1	Sequence 1, Appl
C 32	34	4.9	2853	3	US-09-221-938-1	Sequence 1, Appl
C 33	34	4.9	2653	3	US-08-945-476-7	Sequence 7, Appl
C 34	34	4.9	640681	4	US-09-790-988-1	Sequence 1, Appl
C 35	34	4.9	640681	4	US-09-790-988-1	Sequence 1, Appl
C 36	33.6	4.8	395	4	US-09-894-844-45	Sequence 45, Appl
C 37	33.6	4.8	6156	4	US-10-204-708-60	Sequence 60, Appl
C 38	33.4	4.8	417	4	US-08-559-896B-3	Sequence 3, Appl
C 39	33.4	4.8	417	4	US-09-351-794A-3	Sequence 3, Appl
C 40	33.4	4.8	1956	4	US-08-559-896B-1	Sequence 1, Appl
C 41	33.4	4.8	1956	4	US-09-351-794A-1	Sequence 1, Appl
C 42	33.4	4.8	4557	4	US-09-976-594-592	Sequence 592, Appl
C 43	33.2	4.8	474	4	US-09-621-976-18033	Sequence 18033, A
C 44	33.2	4.8	1696	1	US-08-340-539A-11	Sequence 11, Appl
C 45	33.2	4.8	1696	2	US-08-461-592B-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fis
US-08-232-463-14

Query Match 12.8%; Score 89; DB 1; Length 7218;

[illegible]

```

RESULT 2
US-08-687-580B-6/c
; Sequence 6, Application US/08687580B
; Patent No. 6291647
;
; GENERAL INFORMATION:
;
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: FONSTEIN, Anne Sliene
; APPLICANT: KROON-SWART, Saskia
; APPLICANT: VAN DEVENTER-TROOST, Johanna Petronella Els
; APPLICANT: OHL, Stephan Andres
; APPLICANT: BRES-VLOEWANS, Alexandra Aleida
; APPLICANT: LOGEMANN, Jrgen
; APPLICANT: SELA-BUURLAGE, Marianne Beatrix
;
; TITLE OF INVENTION: Antifungal proteins, DNA coding therefor, and
; TITLE OF INVENTION: hosts incorporating same
;
; NUMBER OF SEQUENCES: 20
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: LADAS & PARRY
;
; STREET: 26 WEST 61 STREET
;
; CITY: NEW YORK
;
; STATE: NY
;
; ZIP: 10023
;
; COUNTRY: U.S.A.
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 3-1/4" Disk 1.44MB
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: Microsoft Windows for Workgroups 3.11
;
; SOFTWARE: Wordperfect 8 for Windows
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/687,580B
;
; FILING DATE: 20-NOV-1996
;
; CLASSIFICATION: 536
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: PCT/EP95/00488
;
; FILING DATE: 2-SEP-1994
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: NL94200321.1
;
; FILING DATE: 9-FEB-1994
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: MASS, Clifford J.
;
; REGISTRATION NUMBER: 30,086

```

REFERENCE/DOCKET NUMBER: U-010895-8

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 708 1890
TELEFAX: (212) 246 8959
TELEX: 233288

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Nicotiana tabacum
STRAIN: Samsun NN
TISSUE TYPE: leaf
IMMEDIATE SOURCE:
CLONE: lambda 3

FEATURE:
NAME/KEY: exon
LOCATION: 774..911
FEATURE:
NAME/KEY: exon
LOCATION: 1045..1154
FEATURE:
NAME/KEY: exon
LOCATION: 3105..3375
FEATURE:
NAME/KEY: exon
LOCATION: 4226..4328
FEATURE:
NAME/KEY: exon
LOCATION: 4432..4777
FEATURE:
NAME/KEY: exon
LOCATION: 4891..5253
FEATURE:
NAME/KEY: exon
LOCATION: 5336..5414
FEATURE:
NAME/KEY: exon
LOCATION: 5531..5743
FEATURE:
NAME/KEY: intron
LOCATION: 912..1044
FEATURE:
NAME/KEY: intron
LOCATION: 1155..3104
FEATURE:
NAME/KEY: intron
LOCATION: 3376..4225
FEATURE:
NAME/KEY: intron
LOCATION: 4329..4431
FEATURE:
NAME/KEY: intron
LOCATION: 4778..4890
FEATURE:
NAME/KEY: intron
LOCATION: 5254..5335
FEATURE:
NAME/KEY: intron
LOCATION: 5415..5530
FEATURE:
NAME/KEY: CDS
LOCATION: join(774..911, 1045..1154, 3105..3375, 4226..4328, 4432..4777, 4891..5253, 5336..5414, 5531..5743)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 774..839
FEATURE:
NAME/KEY: polyA_site
LOCATION: 6095..6100


```

; OTHER INFORMATION: /label= putative
US-08-687-580B-6
Query Match          5.9%; Score 40.8; DB 3; Length 6305;
Best Local Similarity 54.7%; Pred No. 0,065; 67; Indels 0; Gaps 0;
Matches 81; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 72 TTGGAAGACATCAGTCTGAGCAGTAAATGAGATGATTCAGGAAGAGAGTATAAACTGGGA 131
Db 3093 TTCCAAGATGTCAAATTTAAATCTTTTAAATAAGGTAATTCAGAAAATATTGTAATATAGTA 3034
QY 132 ACAGGACGACGACACAGGCTCAGGACACATATTTAAGGACTGGGTAGAAAACCAAGAGAGT 191
Db 3033 GAAGGCGATTGATTATCTATATATTATTAATATTATTAAGATGTAGGTCAGAAATTAAT 2974
QY 192 ATGAACAAGAGTGAGGAGATATTATAGCA 219
Db 2973 AGAAACAAAGAGCTGAAATGCCCAACA 2946

RESULT 3
US-09-334-818A-8/c
; Sequence 8, Application US/09334818A
; Patent No. 6479267
; GENERAL INFORMATION:
; APPLICANT: DAVIS, MARIA
; APPLICANT: FULLER, CARL W.
; APPLICANT: NAMONE, JOSEPH A.
; APPLICANT: HUANG, LIN
; TITLE OF INVENTION: FY7 POLYMERASE
; FILE REFERENCE: PB9817
; CURRENT APPLICATION NUMBER: US/09/334,818A
; PRIOR FILING DATE: 1999-06-17
; CURRENT APPLICATION NUMBER: 60/089,556
; PRIOR FILING DATE: 1998-06-17
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Thermus sp.
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (266)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (293)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (299)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (301)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (304)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (312)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (335)..(338)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (341)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (343)..(346)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (349)..(350)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (352)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (358)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (363)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (366)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (372)..(375)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (379)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (385)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (391)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (393)..(395)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (400)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (403)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (406)..(408)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (422)
; OTHER INFORMATION: a, t, c or g
; FEATURE:

```


OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (153385)..(153385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 5.2%; Score 36.2; DB 4; Length 1664976;
Best Local Similarity 50.9%; Pred. No. 21;

	Matches	86;	Conservative	0;	Mismatches	83;	Indels	0;	Gaps	0;
Qy	476	TTTGGCAACATTCCTCTCTCTGAAATATGAAGAGAGGGGATAAA-TACTGGAGTAGGAT	535							
Db	751136	TTCAAGAAATAAATCCCTCTATGTTTAAAAATATAAATGATATTCATAAATAAAAAATGAAC	751077							
Qy	536	TGTGAAAAAAGTCAAGAGAAAAAAGAACAGACCCCAAGGTAAACAGATACTTCTCCATG	595							
Db	751076	AGATAAAATTTTAAACAGAAAAAGAGAAAAATGCATAATAATCTTTTAAACTCTATATT	751017							
Qy	596	GGATGGTAAAAAGGAAGTTATTTCATAATAATGACCCCTTGGAGGAGT	644							
Db	751016	GGAGGCGATTAACCGGAGTTATAACTAAAGGTTTCTTACTTTTTTAATGGT	750968							

RESULT 12
 US-08-477-831C-1
 ; Sequence 1, Application US/08477831C
 ; Patent No. 6429291
 ; GENERAL INFORMATION:
 ; APPLICANT: TURLEY, EVA A.
 ; APPLICANT: SHUEN, ZHANG
 ; APPLICANT: ENTWISTLE, JOYCELYN
 ; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
 ; NUMBER OF SEQUENCES: 60
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FISH & NEAVE
 ; STREET: 1251 AVENUE OF THE AMERICAS
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: U.S.A.
 ; ZIP: 10020-1104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Rel. #1.0, ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,831C
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PIERRI, MARGARET A.
 ; REGISTRATION NUMBER: 30,709
 ; REFERENCE/DOCKET NUMBER: SIM-10
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-596-9000
 ; TELEFAX: 212-596-9090
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1821 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; DESCRIPTION: /desc = "RHAMM I cDNA (coding region)"
 US-08-477-831C-1

Query Match	5.2%	Score 36;	DB 4;	Length 1821;
Best Local Similarity	45.2%;	Prod. No. 0.91;		
Matches 132;	Conservative 0;	Mismatches 160;	Indels 0;	Gaps 0;
QY	290	ATTITGGGGTGGAGGAAAGGAGGCAATCAAAAGGTGGGGAAAGGCGCTTTTGGG	349	
DB	1117	AATTGACTCTACAAGAAAAAGTAGCTATGGCTGAAAAAAGTGTAGAGATGTTCAACAG	1176	
QY	350	AGGTGGCAGTGTGGGTAGGGAGTAGGGGCTCCGGTTTAGAATACATGGATCAAAAGAAAGTG	409	
DB	1177	CAGATATTGACAGCTGAGAGCACAATCAAGATATGCAAGGATGGTTCAAGATTTGCAG	1236	
QY	410	AGCAGAGCCATTTCCTTTACAACAGAAATTAACAGATTTCCCGAGCTCTTAACCAAGAAATAC	469	
DB	1237	AACAGATCAACCTTTAAAGAAAGAGAAATTTAAAGAAATCACATCTTTCATTTCTTCAGAAA	1296	

Qy 470 AACAGTTTGGNACATCTCTCTCTCAAAATATGAAGAGAGGGGATAAATACATCGGAG 529

Db 1297 ATAACTGATTTGAAAATCACTCAGACAACAGATGAAGACTTTAGGAGCAGACTGGAA 1356

Qy 530 TAGGATTTGTGAAAAAGTCAAGAGAAAAAAAAGAACAGCCCAAGTTGTAA 581

Db 1357 GAGAAAGGAAAAAAGAACAGCAGAGAAAGAAATGTATATCAGAGAAATTAACCC 1408

```

RESULT 13
US-08-477-831C-9
; Sequence 9, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRI, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; DESCRIPTION: /desc = "RHAMM IB"
US-08-477-831C-9

```

	Query Match	5.2%	Score 36	DB 4	Length 1885
	Best Local Similarity	45.2%	Pred. No. 0.92		
	Matches 132	Conservative 0	Mismatches 160	Indels 0	Gaps 0
QY	290	ATTTTGGGGTGAAGGAAAGGAGGCATCAATCAAAGCTGGGGAAAGGCCCTTTTGGG	349		
DB	1191	AATTTGACTCTCAAGAANAAGTASCCTATGGCTGANAAGCTGTAGAAGATGTTTCAACAG	1240		
QY	350	AGGTGGCAGTGTGGTATGGGAGTAGGGGCTCGGGTTAGAAATACATGGATGAAGNAAGGTG	409		
DB	1241	CAGATATTGACAGCTGAGACCAAAATCAAGAATATGCAAGGATGGTTTCAAGATTTCGAG	1300		
QY	410	AGCAGCGCATTTCTTTTACAACAGAAATTAACAGATTTCCAGCGCTCTAACCCAAAGAAAC	469		
DB	1301	AACAGATCAACTTTAAAGAGAGAGAATTAAGAAGATCACATCTTCATTTCTTGAGAAA	1360		
QY	470	AACAAAGTTTGGGAACATTCCTCTCTCTGTAATAATGAAGAGAGGGGNTAATAATCTGGAG	529		
DB	1361	ATACTGATTTTGAAAAATCACTCAGACCAACAGATGAAGACTTTAGGAAGCAGCTGAA	1420		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2004, 23:47:42 ; Search time 184.93 Seconds
(without alignments)
15988.422 Million cell updates/sec

Title: US-10-051-835-14
Perfect score: 696
Sequence: 1 taaatctgtaggcagt.....ttttgaaaatatataaat 696

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04.*

- 1: geneseqn1980s.*
- 2: geneseqn1990s.*
- 3: geneseqn2000s.*
- 4: geneseqn2001as.*
- 5: geneseqn2001bs.*
- 6: geneseqn2002s.*
- 7: geneseqn2003as.*
- 8: geneseqn2003bs.*
- 9: geneseqn2003cs.*
- 10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	681.8	98.0	17947	AAK77679	AAK77679 Human imm
C 2	44.8	6.4	9415	ABL32295	ABL32295 Human imm
C 3	43.8	6.3	2000	ADA71938	ADA71938 Rice gene
C 4	41.8	6.0	300	AAI23067	AAI23067 Probe #13
C 5	41.8	6.0	300	ABA68161	ABA68161 Human foe
C 6	41.8	6.0	300	AAI48373	AAI48373 Probe #17
C 7	41.8	6.0	300	ABA50219	ABA50219 Human bre
C 8	41.8	6.0	300	AAK16541	AAK16541 Human bra
C 9	41.8	6.0	300	ABS41903	ABS41903 Human liv
C 10	41.8	6.0	300	ABS16349	ABS16349 Human gen
C 11	41.8	5.9	451	ABN64891	ABN64891 Human can
C 12	40.8	5.8	6305	AAQ96127	AAQ96127 Tobacco l
C 13	40.6	5.8	6668	ABL33696	ABL33696 Human imm
C 14	40.2	5.8	1269	ACA28673	ACA28673 Prokaryot
C 15	39.8	5.7	32249	ABA17155	ABA17155 Human ner
C 16	39.6	5.7	6319	ABL33643	ABL33643 Human imm
C 17	39.6	5.7	83391	ABQ67094	ABQ67094 Human ang
C 18	39.6	5.7	6681	ABL32155	ABL32155 Human imm
C 19	39.4	5.7	6681	ABL54304	ABL54304 Human imm
C 20	39.4	5.7	46358	ACC42503	ACC42503 Murine P2
C 21	39.4	5.6	9721	AAK65240	AAK65240 Human imm
C 22	38.8	5.6	10048	ABL70314	ABL70314 Chemicall
C 23	38.8	5.6	10048	AA61252	AA61252 Human gen

ALIGNMENTS

RESULT 1

AAK77679/c

ID AAK77679 standard; DNA; 17947 BP.

XX

XX

AC AAK77679;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32491.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180828P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225213P.

QY 1 TAAATATCTGATGCGCGATTAGAAATTTGAGTTTGGAAACACAGAGAGAGCGCTTGTATG 60
 DB 10497 TAAATATCAATAGCGCGATTAGAAATTTGAGTTTGGAAACACAGAGAGAGCGCTTGTATG 10438
 QY 61 GCGATACAGACTTGAAGACATCAGTGTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 120
 DB 10437 GCGATACAGACTTGAAGACATCAGTGTGAGCAGTAAATGAGATGATTCAGGAAAGAGT 10378
 QY 121 ATAACTGGAAGAGGACAGAGGACAGGCTCAAGGAAATATTTAAGCACTGGGTAGAAA 180
 DB 10377 ATAACTGGAAGAGGACAGAGGACAGGCTCAAGGAAATATTTAAGCACTGGGTAGAAA 10318
 QY 181 PACAGAGAGTATGACAGAGTGGAGATTTATGAGTGCACCTTTGAGAGTACATCT 240
 DB 10317 AACAGAGAGTATGACAGAGTGGAGATTTATGAGTGCACCTTTGAGAGTACATCT 10258
 QY 241 CTAGAGTGGTATGTTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCAATTTTGGGGT 300
 DB 10257 CTAGAGTGGTATGTTAGAGCCAGATTTTCAGCATCAGAACCATCAAGCAATTTTGGGGT 10198
 QY 301 GGAAGGAAAGAGCCATCAATCAAGGTGGGAAAGGCGCTTTTGGAGTGGCAGTG 360
 DB 10197 GGAAGGAAAGAGCCATCAATCAAGGTGGGAAAGGCGCTTTTGGAGTGGCAGTG 10138
 QY 361 TGGGTAGGAGTAGGGCTCCGTTAGATACATGATGAAAGAAAGGTGAGCAGCCAT 420
 DB 10137 TGGGTAGGAGTAGGGCTCCGTTAGATACATGATGAAAGAAAGGTGAGCAGCCAT 10078
 QY 421 TTCCTTTACACAGAAATACAGATTTCCAGCCTCTAACCAAGAAACAAAGTTTGG 480
 DB 10077 TTCCTTTACACAGAAATACAGATTTCCAGCCTCTAACCAAGAAACAAAGTTTGG 10018
 QY 481 GAACATTCCTCTCTCTGAAATATCAAGAGAGGGGATAAATACCTGAGTAGGATTGCA 540
 DB 10017 GAACATTCCTCTCTCTGAAATATCAAGAGAGGGGATAAATACCTGAGTAGGATTGCA 9958
 QY 541 AAAAAAGTCAAG-AAAAAAGAACAGCCCAAGTGTAAAGATACCTTCCATGGGAT 599
 DB 9957 AAAAAAGTCAAGAGAAAAAAGAACAGCCCAAGTGTAAAGATACCTTCCATGGGAT 9898
 QY 600 GGTAAAAAGGAAGTTATTTCATATAAATGACCCCTTGAAGAGGTTCAAAAACAGTTGCC 659
 DB 9897 GGTAAAAAGGAAGTTATTTCATATAAATGACCCCTTGAAGAGGTTCAAAAACAGTTGCC 9838
 QY 660 ATATGCTTTATGTGAGTTTGAATAATATATAAAT 696
 DB 9837 ATATGCTTTATGTGAGTTTGAATAATATATAAAT 9801

RESULT 2

ABL32295/c
 ID ABL32295 standard; DNA; 9415 BP.

XX ABL32295;
 AC ABL32295;
 DT 26-MAR-2002 (first entry)
 XX Human immune system associated gene SEQ ID NO: 268.

DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosinetic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

OS Homo sapiens.

XX WO200200928-A2.

PN

XX

PD 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP007537.
 PF 30-JUN-2000; 2000DE-01032529.
 XX 01-SEP-2000; 2000DE-01043826.
 PR (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX Claim 1; SEQ ID NO 268; 32pp + Sequence Listing; German.
 PS The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX Sequence 9415 BP; 1906 A; 297 C; 2611 G; 4601 T; 0 U; 0 Other;
 SQ

Query Match 6.4%; Score 44.8; DB 6; Length 9415;

Best Local Similarity 49.2%; Pred. No. 0.1;

Matches 118; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 387 AATACATGGATGAAGAAAGGTGAGCAGCAGCCATTTCCCTTTACACAGAAATTAACAGATT 446

DB 7617 AATACCGTAAAGAAATATAAATCGTAAAGAAATTTCCCACTAAATAAAAAAA 7558

QY 447 TCCAGCCTTACCAAGAAACACAGTTTGGGAAATCTCTCTCTCTGAAATATGA 506

DB 7557 ACTCCGCTCCAACTAAAAAACAACACGCGTTTAAAAAATACTACCCGATAAAAAAC 7498

QY 507 AAGAGAGGGGATATAATCTAGGATGTTGTAAAAAAGTCAAGAGAAAAAAGAAAC 566

DB 7497 AAAAAACCGAAAAATATAAACCATAAATAAATAAATAAATAAATAAATAAATAA 7438

QY 567 AGCCCAAGTGTAAACAGATCTTCTCCATGGGATGTTAAAAAGGAAGTTATTTCAATAAAA 626

DB 7437 CGACCAAACTAAACAATCTTAATAATAAATAAATAAATAAATAAATAAATAAATAA 7378

RESULT 3

ADA71938
 ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

AC ADA71938;

DT 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

DE Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

XX Oryza sativa.

OS WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

PR

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX Claim 27; SEQ ID NO 5263; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
Query Match 6.3%; Score 43.8; DB 7; Length 2000;
Best Local Similarity 9.9%; Pred. No. 0.11;
Matches 69; Conservative 29; Mismatches 334; Indels 0; Gaps 0;
Qy 2 AATAATCTGATGGCAGTGTAGAAATTTGATTTGGAACACAGAGAGAGGCTTTGATGG 61
Db 307 WRGATWGRGYSRMAATKMYRKYRGKRGWAGRMWRSMCRWSKACYNWRM 366
Qy 62 CGATACAGACTTGAAGACATCAGTCTGAGCAGTAATGAGATGATCAGGAACAGTA 121
Db 367 WRMTTTRWAKSSRTSERRKRWKMRKRYKRYGYSRVRSCRAWRKRCRSGRAWKM 426
Qy 122 TAAACTGGGAGAGGACAGAGGAGGCTCAAGGAACATATTTAAGGACTGGGTAGAAA 181
Db 427 GCRGCMTCRMKS YGMRWKSWKRMASKYKWSRMYRWKWKKSRTTWMGTRGMMGTWG 486
Qy 182 ACAAGACAGTATGACAAAGAGTGGAGATATTTAGCAGTGCCTTTGAGAGTACATCTC 241
Db 487 RCYKKSCKMKRCRRRRGRMYRWKRYMSARYMYRCARKYSYSAARKARCWYRG 546
Qy 242 TAGAGTGGTATGTGTAGAGCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGTG 301
Db 547 KGYWAGWMMKRYKRYMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYK 606
Qy 302 GAAGGAAAGAGGACATGATCAAGAGTGGGGAAGGCTTTGGGAGTGGCAGTGT 361
Db 607 MSWKSMSRSRCKRCKCASKRSARRYAMGGMGTSGSRMSRWKSYTCYWRKWSGMSKSTC 666
Qy 362 GGGTAGGAGTAGGCGCTCCGGTGTAGAATACATGATGAAAGAAAGAGTGACACAGCATT 421
Db 667 WMYWYSKYTYAKYGSYRYRYRAKCMYRWYRYRYRYRYRYRYRYRYRYRYRYRYRY 726
Qy 422 TCCTTTACAGAAATAACAGATTTCCGACCTCTAACCAAGAAACAAACAGATTGGG 481
Db 727 GRYWTSYKYCKSKWSKYSWYWSWYKATWKMRYATRNMMYRYSMKYWTCTW 786
Qy 482 AACATTCTCTCTCTGAATAATCAAGAGAGGGGATAAATCTGGAAGTAGGATTTGAA 541
Db 787 GYWYWRWYKMYKMYKCTKYWYSATYWTGTWAAWMAKTYKMGWMTGAKTGRAR 846
Qy 542 AAAAGTCAAGAGAAAAAAGAACACGCCAAGGTGTAAACAGATCTTCTCCATGGAGTG 601
Db 847 KARYWMAWATKRWTKGAKAWTMAKAWRYKYSWMAWYKYKTRRYKTC 906
Qy 602 TAAAAAGGAGTATTATTCATATAATGACCCCTTGGAGGAGGTCAAAACAGGTTGCCAT 661

Db 907 WVKARWGSWAYRWMMWKSASAKMMWKGGRGKTKYWTCTTWKACGRATKYMCCAGW 966
Qy 662 ATGCTTTATGTGGAGTGTGAAAATATATAATAAAT 696
Db 967 WAMYSYWTRTYMYRTWRWMMWASSRTAKRMWMMW 1001
RESULT 4
AAI23067/c
ID AAI23067 standard; DNA; 300 BP.
XX
AC AAI23067;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13000 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
FN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 13000; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human Hela cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;
Query Match 6.0%; Score 41.8; DB 4; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;
Qy 3 AAATCTCTGATAGGAGTGTAGAAATTTGAGTTTGGAAACACAGAGAGAGGCTTTGATGGC 62
Db 228 AATGTGAGTTTGCAGTTGATTTTGTGCTGGAACCTTAGGAAACATGTTGGGC 169
Qy 63 GA--TACAGACTTGAAGACATCAGTGTCTGAGCAGTAAATGAGATGATTCAGGAAGAGT 120
Db 168 CATCTGCATATCGGTGATGAATGAAGCCACAGAGACGAACATCAATGACGACAGAGT 109

Db 168 CATCTGCATATCGGTGATGAATGAAGCCACAGAAGCAGAGCAATCATCCACCAAGAT 109

QY 121 ATAAACTGGGAAGACGACAGAGGACAGCTCAAGGAA CATATTTAAGGACTCGGTGAAA 180

Db 108 AAATAAAGTGAAGAGAGGGAAGAGGAGGAAGAAAGAGAGAGAGCAATGTTTAAAGG 49

QY 181 AACACAGAGATGATGAACAGAGACTGAGGAGATTATTAGCA 219

Db 48 AATGATAGTGTATTCAAATTTGGCAAGAGACTGAGAA 10

RESULT 9

ABS41903/C

ID ABS41903 standard; DNA; 300 BP.

AC ABS41903;

XX XX

XX 25-FEB-2003 (first entry)

XX Human liver single exon probe, SEQ ID NO 16893.

XX Human; single exon nucleic acid probe; liver; cirrhosis;

KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;

KW coronary heart disease; ss.

XX XX

XX Homo sapiens.

OS OS

PN WO200157273-A2.

XX XX

XX 09-AUG-2001.

XX XX

XX 30-JAN-2001; 2001WO-US000664.

XX XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-SEP-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX XX

FA (MOLE-) MOLECULAR DYNAMICS INC.

XX XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX

XX WPI; 2001-488898/53.

XX XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human adult liver.

XX XX

XX Claim 4; SEQ ID NO 16893; 658pp; English.

XX XX

CC The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult liver.

CC (I) may be used for predicting, measuring and displaying gene expression

CC in samples derived from human adult liver. The genes identified may be

CC involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is

CC associated with coronary heart disease. ABS25011-ABS51005 represent human

CC liver single exon nucleic acid probes of the invention. Note: The

CC sequence information for this patent does not appear in the printed

CC specification but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX XX

SQ Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

QY	3	AAATATCTGATGGCAGTTAGRAATTTGAGTTTGGAAACACAGGAGAGAGCGCTTTGATGGC	62
Db	228	AAATGTGAAGTTTGCAGTTTCAATATTTTGTGCGCTTGGAACTTTAGAAAAACAATGTGTGGSC	169
QY	63	GA--TACAGACTTTGGAAGACATCAGTCTCTGAGCAGTAAATAGAGATGTTACGAAAGAGT	120
Db	168	CATCTGCAATATCGTGATGAATGAAGCCACAGAAAGCAGAAACCAATCAATGCAGCAAGT	109
QY	121	ATTAACCTGGGAAGAGCAGACAGACAGCGCTCAAGGAAACATATTTAAGGACTTGGGTAGAAA	180
Db	108	AAATTAAGTGAAGAGAAAGGAGAGGAGAGAGAAAGAGAGAGAGCAATGTTTAAGG	49
QY	181	AAACAAGAGACTATGAACAAGAGTGCAGGAGATTTATTAGCA	219
Db	48	AATGATAGTGTATTTCAAATTTGGCAAAAGAGACTGAGAA	10
RESULT 10			
ABS16349/c			
XX	ID	ABS16349 standard; DNA; 300 BP.	
XX	AC	ABS16349;	
XX	DT	19-AUG-2002 (first entry)	
XX	DE	Human genome-derived single exon probe ORF from lung SEQ ID NO 16340.	
KW	Human; ds; single exon probe; asthma; lung cancer; COPD; IID;		
KW	chronic obstructive pulmonary disease; interstitial lung disease;		
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;		
KW	tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;		
KW	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;		
KW	pulmonary histiocytosis; lymphangioliomyomatosis; Karagener syndrome;		
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;		
KW	primary ciliary dyskinesia; pulmonary hypertension;		
KW	hyaline membrane disease; open reading frame; ORF.		
XX	Homo sapiens.		
OS	WO200186003-A2.		
XX	PN	15-NOV-2001.	
XX	PD		
XX	PF	04-JAN-2001; 2001WO-US0000565.	
XX	PR	04-FEB-2000; 2000US-0180312P.	
PR	26-MAY-2000; 2000US-0207456P.		
PR	30-JUN-2000; 2000US-00608408.		
PR	03-AUG-2000; 2000US-00632366.		
PR	21-SEP-2000; 2000US-0234687P.		
PR	27-SEP-2000; 2000US-0236359P.		
PR	04-OCT-2000; 2000GB-00024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
PI	WPI; 2002-114183/15.		
XX	Spatially-addressable set of single exon nucleic acid probes, used to		
PT	measure gene expression in human lung samples.		
XX	Claim 4; SEQ ID NO 16340; 634pp; English.		
XX	The invention relates to a spatially-addressable set of single exon		
CC	nucleic acid probes for measuring gene expression in a sample derived		
CC	from human lung comprising single exon nucleic acid probes having one of		
CC	12614 nucleic acid sequences mentioned in the specification, or their		
CC	complements or the 12387 open reading frames derived from the 12614		
CC	probes. Also included are a microarray comprising the novel set of probes		
CC	; the novel set of probes which hybridise at high stringency to a nucleic		
CC	acid expressed in the human lung; measuring gene expression in a sample		

CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Heremansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a single exon probe open reading frame of the
CC invention. Note: the sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 300 BP; 68 A; 87 C; 37 G; 108 T; 0 U; 0 Other;

Query Match 6.0%; Score 41.8; DB 6; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.19;
Matches 115; Conservative 0; Mismatches 102; Indels 2; Gaps 1;
QY 3 AAATATCTGATAGCAGTGTAGAATTTGAGTTTGAACACAGAGAGAGGCTTTGATGCG 62
DB 228 AAATGTGAGTTTGCAGTTGCAATTTTGTGCTTGAACCTTAGGAAAACAATGTGTGGGC 169
QY 63 GA--TACAGACTTGGAGACATCAGTGTCTGAGCAGTAAATGAGATGATTCAGGAAGAGT 120
DB 168 CATCTGCAATTCGGTGAATGAATGAAGCCACAGAGCAGAGCAATCAATCCAGCAGAGT 109
QY 121 ATAACTGGAAGAGGACAGACAGCAGCTCAAGGAACATATTTAAGGACTGGTAGAAA 180
DB 108 AATTAAGTGAAGAGGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 49
QY 181 AACAGAGAGTATGAACAGAGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
DB 48 AATGATAGTGTATTCAAAATTTGGCAAAGAGAGACTGAGAA 10

RESULT 11
ABN64891
ID ABN64891 standard; cDNA; 451 BP.
XX
AC ABN64891;
XX
AC
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 4858.
XX
XX Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200214500-A2.
XX
XX 21-FEB-2002.
PD
XX

PF 16-AUG-2001; 2001WO-US025840.
XX
PR 16-AUG-2000; 2000US-0226326P.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
PI WPI; 2002-241905/29.
DR
XX New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth.
XX
PS Claim 1; SEQ ID NO 4858; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 451 BP; 129 A; 104 C; 106 G; 112 T; 0 U; 0 Other;
Query Match 5.9%; Score 41; DB 6; Length 451;
Best Local Similarity 61.9%; Pred. No. 0.37;
Matches 65; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 10 TGATAGCAGTGTAGAATTTGAGTTTGAACACAGAGAGAGGCTTTGATGGCAGATACAG 69
DB 72 TGTGAGCAGATAGATCTATGAGTTGGAGCTCAGGGGCGAAGTCTGGCTGAGATATAA 131
QY 70 ACTTGGAGACATCAGTGTCTGAGCAGTAAATGAGATGATTCAGGA 114
DB 132 ATTGGAGTCTCAGCAGACAGATGATATTTAAAGCCTGCAGA 176
RESULT 12
AAQ96127/c
ID AAQ96127 standard; DNA; 6305 BP.
XX
AC AAQ96127;
XX
DT 07-MAR-1996 (first entry)
XX
DE Tobacco leaf antifungal protein genomic DNA.
XX
KW Tobacco; plant; fungicide; pesticide; antifungal protein; Phytophthora;
KW transgenic plant; crop improvement; ds.
XX
OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT CDS 774..5743
FT exon /*tag= r
FT 774..911
FT /*tag= a
FT sig_peptide 774..839
FT /*tag= p
FT intron 912..1044
FT /*tag= i
FT exon 1045..1154
FT /*tag= b
FT intron 1155..3104
FT /*tag= j
FT exon 3105..3375


```

FT intron      /*tag= c
FT 3376..4225
FT /*tag= k
FT 4226..4328
FT /*tag= d
FT intron      /*tag= l
FT 4329..4431
FT /*tag= l
FT exon        /*tag= e
FT 4432..4777
FT intron      /*tag= m
FT 4778..4890
FT /*tag= m
FT exon        /*tag= f
FT 4891..5253
FT /*tag= f
FT intron      /*tag= n
FT 5254..5335
FT /*tag= n
FT exon        /*tag= g
FT 5336..5414
FT /*tag= g
FT intron      /*tag= o
FT 5415..5530
FT /*tag= o
FT exon        /*tag= h
FT 5531..5743
FT /*tag= h
FT polyA_signal /*tag= q
FT 6095..6100
FT /*tag= q
FT /*label= putative
XX
XX WO9521929-A1.
PN
XX
XX 17-AUG-1995.
XX
XX 09-FEB-1995; 95WO-EP000488.
XX
XX 09-FEB-1994; 94EP-00200321.
XX
XX (MOGE-) MOGEN INT NV.
XX
XX Melchers LS, Ponstein AS, Kroonswart S, Van Deventer- Troost JPE;
XX Ohl SA, Bres-Vloemans AA, Logemann J, Sela-Buurlage MB;
XX
XX WPI; 1995-293130/38.
XX P-FSDS; AAR78679.
XX
XX New antifungal protein and DNA obtd. from plants - used for producing
XX transgenic plants or in antifungal compsns., partic. for retarding growth
XX of Phytophthora infestans.
XX
XX Disclosure; Page 27; 58pp; English.
XX
XX This DNA sequence may be expressed a part of a chimeric DNA which is
XX either expressed in a host cell or transgenic plant using binary vector
XX , plasmid pMOG846, or a replicon, preferably pMOG841. The resulting
XX expressed protein displays fungicidal activity against Phytophthora
XX infestans. The DNA may be expressed in a tissue- specific manner,
XX especially in the style and stigma
XX
XX Sequence 6305 BP; 1905 A; 1094 C; 1114 G; 2192 T; 0 U; 0 Other;
XX
XX Query Match      5.9%; Score 40.8; DB 2; Length 6305;
XX Best Local Similarity 54.7%; Pred. No. 1.1;
XX Matches 81; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
XX
XX 72 TTGGAACACATCAGTGTGTCAGCAGTAATAGATGATTCAGGAAGAGTATAACTGGGA 131
XX
XX 3093 TTCCAGATGTCAATTTAAATCTTTAAATAAGGTAATTCAGAAATAATTCGTATATAGTA 3034
XX
XX 132 AGAGCAGAGGACAGCGCTCAGGAACATATTTAAGGACTGGGTAGAAAAACAAGAGAGT 191
XX
XX 3033 GAAGGCGATTGATTATCTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2974
XX
XX 192 ATGAACAAGAGTGAGGAGATTATTAGCA 219
XX
XX 2973 AGAAACAAAAGAGCTGAAATGCCCAACA 2946

```

```

RESULT 13
ABL33696/C
ID ABL33696 standard; DNA; 6668 BP.
XX
XX ABL33696;
AC
XX 26-MAR-2002 (first entry)
DT
XX
XX Human immune system associated gene SEQ ID NO: 1669.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytotatic; phototropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1669; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6668 BP; 1628 A; 328 C; 1974 G; 2737 T; 0 U; 1 Other;
XX
XX Query Match      5.8%; Score 40.6; DB 6; Length 6668;
XX Best Local Similarity 50.3%; Pred. No. 1.3;
XX Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
XX
XX 428 ACAACAGAAAATAACAGATTCCCGCCTTAACCAAGAAACACACAGTTGGGAACATT 487
XX
XX 5498 ACAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5429
XX
XX 498 CCTCTCTTCTGAAATATCAAGAGAGGGGATAAATACTGGAGTAGGATTGTGAAAAAAGT 547
XX
XX 5428 CTTACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5369
XX
XX 548 CAAGAGAAAAAATAAAGAACAGCCCAAGTGTAAACAGATACCTTCTCATGGGATGGTAAAAA 607
XX
XX 5368 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5309
XX
XX 608 GGAAGTTATTCATAAAA 526
XX
XX 5308 CAANAATTTTCTAAAAAAA 5290

```

RESULT 14
ACA28673 standard; DNA; 1269 BP.
XX ID ACA28673 standard; DNA; 1269 BP.
XX AC ACA28673;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #10330.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Clostridium botulinum.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX PR WPI; 2003-029926/02.
XX PR P-PSDB; ABU24803.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 16543; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
XX CC prokaryotic essential genes. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1269 BP; 498 A; 116 C; 217 G; 438 T; 0 U; 0 Other;
Query Match 5.8%; Score 40.2; DB 7; Length 1269;
Best Local Similarity 47.6%; Pred. No. 0.91;
Matches 151; Conservative 0; Mismatches 163; Indels 3; Gaps 1;
QY 19 GTTAGAATTTGAGTTTGGACACAGGAGAGGCTTTGATGGCGATACAGACTTGAAG 78
DB 400 GCITTAATATAGGCTATATAACATTCAGTGGAGACTTTATAGCCGTATATGATGCAGAT 459
QY 79 ACATCAGTCTGAGCAGTAAATGAGATGATTCAGGAAAGAGTATAAATCTGGGAGAGGAC 138
DB 460 AATACTCCAGATAAAATGCATTAAAATATCTTATGAAAATCTATAATTTGAGGACGAGGT 519
QY 139 AGAGCAGAGCTCAAGGACACATATTTAAGGACTGGGTAGAAAACAGAGAGATATGAACA 198
DB 520 TTAGGAGCAGTTATAGGAA---AGTTTAGAATAGAAATTAAGACAGAAATATGTTTAA 576
QY 199 AGAGTGAGGAGATTTATAGCAGTGCACCTTTGAGAGTACATCTCTAGAGTGGTATGTTAG 258
DB 577 AGATTTATAATATAGAGACCTTAAGTTTCAATGGATGTGCCAAGCAGGAGATGAAG 636
QY 259 AGCCAGATTTTCAGCATCAGAACCATCAAGCATTTTGGGGGTGAGGAAAGAGGAGCCAT 318
DB 637 TTATTAATTTTATGTACTATACCTGGCACAATTTTGTAGTAAAGAAATATAATACAA 696
QY 319 GAATCAAAAGGTGGGA 335
DB 697 GAGTTGATGATGGGA 713
RESULT 15
ABAL17155/c
ID ABAL17155 standard; DNA; 32249 BP.
XX AC ABAL17155;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 9486.
XX KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
XX KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX KW antiparkinsonian; antiscikling; antianaemic; antiarthritic; cancer;
XX KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX PN WO200159063-A2.
XX PD 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001334.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225472P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231142P.
PR 08-SEP-2000; 2000US-0231143P.
PR 08-SEP-2000; 2000US-0231144P.
PR 08-SEP-2000; 2000US-0231143P.
PR 08-SEP-2000; 2000US-0231141P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237057P.
PR 02-OCT-2000; 2000US-0237058P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249259P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-541565/60.
Nucleic acids encoding 324 human nervous system antigen polypeptides,
useful for preventing, diagnosing and/or treating nervous system cancers
and metastases.
Disclosure; SEQ ID NO 9486; 1701pp + Sequence Listing; English.
The invention relates to novel genes (ABAI1004-ABA21534) and proteins
(ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
medical conditions e.g. by protein or gene therapy. The genes are
isolated from a range of human tissues disclosed in the specification.
The nucleic acids, proteins, antibodies and (ant)agonists are useful in
the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
disorders e.g. Addison's disease, allergies, autoimmune haemolytic
anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
cardiovascular disorders such as myocardial ischaemias; (d) wound healing

